Towards a data model for (inter)textual relationships

Connecting Ancient Egyptian texts and understanding scribal practices

Stéphane Polis (F.R.S.-FNRS – ULg)
Vincent Razanajao (ULg)
Nathalie Sojic (ULg)
Outline of the talk

- Background information about the Ramses Project & Ramses Online
- Evolution of the data model between 2006 and 2016
- Data models: state-of-the-art in digital editing
- Towards a data model for (inter)textual relationships in Ancient Egyptian
The Ramses Project & Ramses Online
Background information
The Ramses Project

Goal
- Build a richly annotated corpus of Late Egyptian texts
The Ramses Project

Goal

- Build a richly annotated corpus of Late Egyptian texts
The Ramses Project

Goal
- Build a richly annotated corpus of Late Egyptian texts
- Useful both for philologists and linguists
The Ramses Project

JAVA software (MySQL – texts stored in XML)

- LexiconEditor
- TextEditor
The Ramses Project

What kind of data?
- Hieroglyphic spellings
The Ramses Project

What kind of data?
- Hieroglyphic spellings
- Lemmatization and morphological annotation
The Ramses Project

What kind of data?

- Hieroglyphic spellings
- Lemmatization and morphological annotation
- Textual criticism
The Ramses Project

What kind of data?
- Hieroglyphic spellings
- Lemmatization and morphological annotation
- Textual criticism
- Translation (French / English)
The Ramses Project

History (2006-2016)
The Ramses Project

The corpus

- Number of witnesses

![Graph showing the increase in the number of witnesses over time, with a significant increase starting from 2010.](image)
The Ramses Project

The corpus
- Number of occurrences

The graph shows the number of occurrences from 2006 to 2015, with a steady increase each year.
Ramses Online (ramses.ulg.ac.be)

Ramses Online un corpus annoté du néo-égyptien

Recherche simple dans les textes

<table>
<thead>
<tr>
<th>Lemmes</th>
<th>Graphies</th>
<th>Traductions</th>
</tr>
</thead>
</table>

Translittération ou traduction

**Filtrer sur**
- Translittération
- Traduction
- Les deux

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>mais, vraiment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ITCL)</td>
</tr>
<tr>
<td>3.t</td>
<td></td>
<td>moment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SUBST)</td>
</tr>
<tr>
<td>3vv</td>
<td></td>
<td>enduire, crêpir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(VB)</td>
</tr>
<tr>
<td>3vc</td>
<td></td>
<td>faire du tort</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(VB)</td>
</tr>
</tbody>
</table>
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
- With powerful linguistic searching capabilities
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
- With powerful linguistic searching capabilities
- In interaction with its users
Ramses Online (ramses.ulg.ac.be)

- Responsive website based on Bootstrap
- With powerful linguistic searching capabilities
- In interaction with its users

3e mois de Péret, 10e jour :
Collaborations

- TEI interchange format
  - Laurent Coulon (EPHE – Paris)
  - Frederik Elwert (CERES – Bochum)
  - Emmanuelle Morlock (HiSoMA – CNRS)
  - Stéphane Polis (F.R.S.-FNRS – Liège)
  - Vincent Razanajao (ULg – Liège)
  - Serge Rosmorduc (CNAM – Paris)
  - Simon Schweitzer (BBAW – Berlin)
  - Daniel A. Werning (EXC Topoi – Berlin)
Collaborations

- **TEI interchange format**
  - Laurent Coulon (EPHE – Paris)
  - Frederik Elwert (CERES – Bochum)
  - Emmanuelle Morlock (HiSoMA – CNRS)
  - Stéphane Polis (F.R.S.-FNRS – Liège)
  - Vincent Razanajao (ULg – Liège)
  - Serge Rosmorduc (CNAM – Paris)
  - Simon Schweitzer (BBAW – Berlin)
  - Daniel A. Werning (EXC Topoi – Berlin)

- **Linked metadata**
  - Thot (thot.philo.ulg.ac.be)
Collaborations

- TEI interchange format
  - Laurent Coulon (EPHE – Paris)
  - Frederik Elwert (CERES – Bochum)
  - Emmanuelle Morlock (HiSoMA – CNRS)
  - Stéphane Polis (F.R.S.-FNRS – Liège)
  - Vincent Razanajao (ULg – Liège)
  - Serge Rosmorduc (CNAM – Paris)
  - Simon Schweitzer (BBAW – Berlin)
  - Daniel A. Werning (EXC Topoi – Berlin)

- Linked metadata
  - Thot (thot.philo.ulg.ac.be)
  - Trismegistos (http://trismegistos.org/)
Collaborations

- TEI interchange format
  - Laurent Coulon (EPHE – Paris)
  - Frederik Elwert (CERES – Bochum)
  - Emmanuelle Morlock (HiSoMA – CNRS)
  - Stéphane Polis (F.R.S.-FNRS – Liège)
  - Vincent Razanajao (ULg – Liège)
  - Serge Rosmorduc (CNAM – Paris)
  - Simon Schweitzer (BBAW – Berlin)
  - Daniel A. Werning (EXC Topoi – Berlin)

- Linked metadata
  - Thot (thot.philo.ulg.ac.be)
  - Trismegistos (http://trismegistos.org/)
  - TopBib (http://topbib.griffith.ox.ac.uk/)
Evolution of the data model

Between 2006 and 2016
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016)
- Towards TDM 2.0
Evolution of the data model

- The original model (2006)

Following the egyptological practice, the decision was made to encode in hieroglyphs (and to annotate) every single witness of each text (envisioned as an abstraction). The document, on the other hand, was seen as the object on which multiple witnesses could occur.

⇒ No ideal/intended text, rather individual scribes as agents
Evolution of the data model

- The original model (2006)

Synoptic edition of *Sinuhe* (Koch 1990: 8,2-7)
Evolution of the data model

- The original model (2006)

Synoptic edition of *Sinuhe* (Koch 1990: 8,2-7)
Evolution of the data model

- The original model (2006)

Synoptic edition of *Sinuhe* (Koch 1990: 8,2-7)

Witness 1
Witness 2
Witness 3
Witness 4
Witness 5
Witness 6

One text
Evolution of the data model

- The original model (2006)
Evolution of the data model

- The original model (2006)
  - Witness 1
    - Literary text (hymn)
  - Witness 2
    - Administrative text (account)

One document
Evolution of the data model

- The original model (2006)

  Witness 1
  Literary text (hymn)

  Witness 2
  Administrative text (account)

One document
Evolution of the data model

- The original model (2006)

“The [documentary and textual] dimensions are incapable of disconnection: (...) they negatively constitute one another and (...) this constitution requires human agency at every step, from composition to reception.”

(Eggert 2004: 169; see already Eggert 1998: 57)
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology

<table>
<thead>
<tr>
<th>Object(s)</th>
<th>Document</th>
<th>Witness(es)</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>O. IFAO 1255 A (v°) + O. IFAO 1255 B (v°) + (O. Varille 39 [= V]; lost)</td>
<td>Reconstructed original potsherd ostracon based on the three Objects</td>
<td>Occurrence of the <em>Teaching of Amennakhte</em> (§39-48) on O. IFAO 1255 B (v°) [+ V lost]</td>
<td>Abstract version of the <em>Teaching of Amennakhte</em>, viewed here as a Synoptic edition</td>
</tr>
</tbody>
</table>
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology

<table>
<thead>
<tr>
<th>Object(s)</th>
<th>Document</th>
<th>Witness(es)</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. IFAO 1255 A (v^o) + O. IFAO 1255 B (v^o) + (O. Varille 39 ([- V); lost)</td>
<td>Occurrence of the Teaching of Amennakhte (\S 39-48) on O. IFAO 1255 B (v^o) ([+ V lost]</td>
<td>Abstract version of the Teaching of Amennakhte, viewed here as a Synoptic edition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reconstructed original potsherd ostracron based on the three Objects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
- Towards TDM 2.0
  - Why?
Evolution of the data model

*P. Mag. Isis* (P. CGT 54051) = the main Witness of the Text (*Isis magical papyrus*)

Other Witnesses of the Text, occurring on different Documents (ostraca)

O. Gardiner  
O. Nash 14  
O. DeM 1048  
O. DeM 1046
Evolution of the data model

P. Mag.Isis (P. CGT 54051 r°)

P. Chester Beatty XI (r°)

P. CGT 54051 is composed of
- A group of incantations (r° 2,1–8)
- A formula for bandaging (r° 2,8–12)
- The story of Isis & Ra (r° 2,12 – 5,5)
- A group of formulae (r° 5,12 – 6,3)
- A group of formulae (v° 2,6 – 5,1)

P. Chester Beatty XI is composed of
- The story of Isis & Ra (r° 1,1 – 4,1)
- A group of magical spells (r° 4,2–10)
- A magical text against scorpions (r°, fgmts A–L)
- The end of a magical text (v° 1,1–3)
- An account (v° 1,4–10)
- An hymn to the god Amun (v° 2–3)
- A group of formulae for “safety upon the river” (v°, fgmts A–D)
Evolution of the data model

Witness A = « P. Mag. Isis »

Section 1

Section 2

Section 3

Witness B on « P. Chester Beatty XI »

Section 1

Section 2

Section 3
Evolution of the data model

Witness A = « P. Mag. Isis »

Section 1

Section 2

Section 3

Text « Isis and Ra »

Witness B on « P. Chester Beatty XI »

Section 1

Section 2

Section 3

hasPart

isActualizationOf

hasPart

isActualizationOf

hasPart

hasPart
Evolution of the data model

Witness A = « P. Mag. Isis »

Section 1

Section 2

Section 3

Witness B on « P. Chester Beatty XI »

Section 1

Section 2

Section 3

Text « Isis and Ra »

Complex Text « Magical - Isis »

Complex Text « Magical - Beatty »
Evolution of the data model

Witness A = « P. Mag. Isis »

Text « Isis and Ra »

Witness B on « P. Chester Beatty XI »

Complex Text « Magical - Isis »

Complex Text « Magical - Beatty »
Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016)
- Towards TDM 2.0
  - Why?
  - With which model can we handle such cases?
    - Relationships between witnesses
    - Relationships between texts
Data models

State-of-the-art in digital editing
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (NLP)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

**text-pair**

PAIR: Pairwise Alignment for Intertextual Relations

PAIR (Pairwise Alignment for Intertextual Relations) is a simultaneous implementation of a sequence alignment algorithm for human text analysis designed to identify "similar passages" in large collections of texts. These may include direct quotations, plagiarism and other forms of borrowings, commonplace expressions and the like. We are developing two currently streams:

**PhilolLine** (for PhiloLogic alignment) is the experimental model, written largely by Mark (in very old fashioned perl), designed to perform all-against-all comparisons in documents loaded into a PhiloLogic database. An entire corpus is indexed and compared against itself, or another database, to find text reuse.

March 2009: We have released version 0e of PhiloLine. To get some sense of what this does and how it currently works, please consult our ARTFL-Frantext Release Notes. We have put this into production for ARTFL and are using it for a number of other projects. We have fairly extensive installation and use instructions in our Release Notes. This is known to work for English, French, Greek, and other language document collections up to 11,500 documents and 520 million words.
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework

The text is seen as “a mosaic of quotations; any text is the absorption and transformation of another” (1986: 37)
Data models for digital editing

- 'Digital approaches to intertextuality' is a hot topic

- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework

“any text is a new tissue of past citations. Bits of code, formulae, rhythmic models, fragments of social languages, etc., pass into the text and are redistributed within it, for there is always language before and around the text” (1981: 39)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic

- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic

- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework
  G. Genette (1997 [1982]), *Palimpsests: Literature in the second degree*
  - Intertextuality
  - Paratextuality
  - Metatextuality
  - Hypertextuality
  - Architextuality
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic

- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework

- Data models
  - “In the context of digital editing, by modelling we mean at least two types of conceptualization: the one that tries to organize entities such as texts, documents, works, along with their relationships and how they have happened to come into being, and the analytical process of establishing the kind and purpose for the production of a new edition, its implied community of users and what features best represent their various needs.” (Pierazzo 2015: 48)
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
  - The FRBR data model (1997 [2009])
Data models for digital editing

- Four basic elements

![Diagram showing the relationship between Work, Expression, Manifestation, and Item]
Data models for digital editing

- Four basic elements
- Relationships between these elements

- $w_1$ Robertson Davies’ *The Deptford trilogy*
  - $w_{1.1}$ Robertson Davies’ *Fifth business*
  - $w_{1.2}$ Robertson Davies’ *The manticore*
  - $w_{1.3}$ Robertson Davies’ *World of wonders*
Data models for digital editing

- Four basic elements
- Relationships between these elements
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
  - The FRBR data model (1997 [2009])
  - ‘Neo-platonistic idealism’
  - Cf. Rafferty (2015) “the notion of work points to intertextuality, with all its potential for rich analysis, but at the same time it embeds deep in its system the logocentrism of the ideal signified”

FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS

Final Report

IFLA Study Group on the Functional Requirements for Bibliographic Records

Approved by the Standing Committee of the IFLA Section on Cataloguing

September 1997
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
  - The FRBR data model (1997 [2009])
  - FRBRoo (> CIDOC-CRM)
Data models for digital editing

- FRBRoo is the FRBR ontology revised and expressed in an object-oriented form compatible with that of the CIDOC-CRM
- Strong orientation towards modelling intellectual processes
- FRBRoo is the FRBR ontology revised and expressed in an object-oriented form compatible with that of the CIDOC-CRM
- Strong orientation towards modelling intellectual processes
- Addition of some interesting elements

Data models for digital editing
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
  - The FRBR data model (1997 [2009])
  - FRBRoo (> CIDOC-CRM)
  - Pierazzo (2015), *Digital Scholarly Editing*
Data models for digital editing

- Pierazzo (2015)
  - Interpretative process
  - Multiple dimensions
  - A single occurrence of ‘intertextuality’ in the book, about the documents’ “literary dimension: style, rhetorical features, genre, intertextuality, citations and allusions.”

Figure 2.1: Conceptual model of texts and documents
Data models for digital editing

Pierazzo (2015)
Data models for digital editing
Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic

- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework

- Data models
  - The FRBR data model (1997 [2009])
  - FRBRoo (> CIDOC-CRM)
  - Pierazzo (2015), Digital Scholarly Editing
Data models for digital editing

- Hedges et al. (2016, in press) – SAWS
- They worked with an extension of FRBRoo
Data models for digital editing

- Hedges et al. (2016, in press) – SAWS
- They worked with an extension of FRBRoo

All the ‘intertextual’ relationships are defined at the level of E33 ‘Linguistic Object’
‘Digital approaches to intertextuality’ is a hot topic

Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

Theoretical framework

Data models
- The FRBR data model (1997 [2009])
- FRBRoo (> CIDOC-CRM)
- Pierazzo (2015), Digital Scholarly Editing
- Hedges et al. (2016, in press)
Data models for digital editing

- To sum up
  - The concept *work* as an abstract/unifying element represented by witnesses/expressions
Data models for digital editing

To sum up

- The concept *work* as an abstract/unifying element represented by witnesses/expressions
- ‘Part-whole’ relationships
  - Work-level: ‘Complex’ works made of ‘simpler’ works
  - Witness-level: ‘Complex’ witnesses made of smaller witnesses
Data models for digital editing

- To sum up
  - The concept work as an abstract/unifying element represented by witnesses/expressions
  - ‘Part-whole’ relationships
    - Work-level: ‘Complex’ works made of ‘simpler’ works
    - Witness-level: ‘Complex’ witnesses made of smaller witnesses
  - Other relationships
    - Between different works (‘sequel’, ‘imitation’, etc.)
    - Between different witnesses (‘verbatim copy of’, ‘translation’, etc.)
Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*

- **Texts of documents**, namely the texts one can derive from physical documents

- **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)
Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*
  - **Texts of documents**, namely the texts one can derive from physical documents
  - **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)
Thomas Tanselle (1989), *A Rationale of Textual Criticism*

- **Texts of documents**, namely the texts one can derive from physical documents
- **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)
Thomas Tanselle (1989), *A Rationale of Textual Criticism*

- **Texts of documents**, namely the texts one can derive from physical documents
- **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)

\[ \text{Witness} \quad \xrightarrow{\text{Content}} \quad \xleftarrow{\text{Expression}} \quad = \text{Tanselle’s ‘Text of the document’} \]

\[ = \text{Hjelmslev’ basic distinction} \]
Thot Data Model (2.0)

- The Domain and Range of intertextual relationships become much clearer when this distinction is made (cf. Büchler’s opposition between ‘syntactic and semantic text reuses’).
The Domain and Range of intertextual relationships become much clearer when this distinction is made (cf. Büchler’s opposition between ‘syntactic and semantic text reuses’).

An analysis of the relationships suggested in the literature suggests that they can be organized according to three types for any element of the model.

- Syntagmatic
- Paradigmatic
- Whole/part
Thot Data Model (2.0)

- The Work

*Diagram showing relationships between Work, isComposedOf/isComponentOf, isSuccessorOf/hasSuccessor, isSequelTo, isDerivativeOf/hasDerivative, isSummarizationOf, isAdaptationOf, isTransformationOf, and isImitationOf.*
• The Work
• The Witness
  • Expression
  • Content

Thot Data Model (2.0)
Thot Data Model (2.0)

- General model
Thot Data Model (2.0)

- General model
Thot Data Model (2.0)

Illustration 1. Multilingual decree

Decree of Canopus by Ptolemy III

1. Stela Cairo CG 22186
   Decree of Canopus, Copy from ancient Imau

2. Stela Louvre C 122
   Decree of Canopus, Copy from ancient Memphis (?)

3. Stela Port-Said Museum Inv. 493
   Fragment of the Canopus Decree, Copy from Bubastis

+ Stela Cairo CG 22187
+ Fragment from El-Kab

Full extent of reconstructed text

Lines 2-9 of canonical text (= version 1)

Decree of Ptolemy III Euergetes
This papyrus is several centuries ‘older’ than P. CGT 54051. It bears a short hymnal sequence that also occurs in P. Mag. Isis

Roccati (2011: 141)
Thot Data Model (2.0)
Implementation
Thanks!