

Laurent Romary and Zina Tucsnak INRIA-LORIA, CNRS-ATILF LREC Symposium: The Open Language Archives Community 29 May 2002

ATILF Archives (1)

ATILF's computerized linguistic resources for lexical and textual analysis in French language

- FRANTEXT: A French textual database
 - 3417 written texts: French literature from 19th to 20th
 - 1940 texts annotated with Part-of-Speech tags

OLAC Launch, LREC-02

ATILF Archives (2)

- DICTIONARIES and ENCYCLOPEDIAS
 - TLFi (Trésor de la Langue Française Informatisé) Computerized dictionary containing 100,000 head words, 270,000 definitions and 300,000 examples
 - DIDEROT and D' ALEMBERT Encyclopedia 72,000 articles written by more than 140 contributors, 20.8 million words, 2569 plates

DLAC Launch, LREC-02

ATILF Archives (3)

- **DICTIONARIES and ENCYCLOPEDIAS**
 - Académie française's Dictionaries
 - 1st edition (1694), 5th edition (1798), 6th edition (1835), 8th edition (1932-1940), 9th edition
 - Old Dictionaries
 - Robert Estienne's Dictionarium latinogallicum (1552)
 - Jean Nicot's Thresor de la langue françoise (1606)
 - Pierre Bayle's Dictionnaire historique et critique (1740)

Why OLAC for ATILF? (1)

- Current end-users of ATILF resources: a small community
 - Free access for TLFi and the other dictionaries
 - Annual subscription for Frantext and Diderot and d' Alembert's encyclopedia (actually 200 subscribers

OLAC Launch, LREC-02

Why OLAC for ATILF? (2)

- Connections number (daily researches not only hits)
 - 550 for *TLFi*
 - 450 for Académie Française's Dictionaries
 - 400 for Old Dictionaries
 - 250 for FRANTEXT and 150 for Diderot and D' Alembert
- By involving in OLAC , ATILF resources can be found, used, cited

OLAC Launch, LREC-02

ATILF: VIDA Resource Creator

- ATILF joined OLAC as a Virtual Data Provider
 - Frantext archive
 - Bibliography of the entire corpus
 - Dictionaries and Encyclopedia archive
 - TLFi,Old Dictionaries,Académie Française's Dictionaries and Diderot and d' Alembert Encyclopedia

OLAC Launch, LREC-02

ATILF Metadata

- A wide variety of resources on different platforms, using proprietary metadata format
 - Frantext,TLFi and two of the Académie Française's Dictionaries run on Windows OS with the software "STELLA"
 - Old Dictionaries, Diderot and d' Alembert Encyclopedia and three of the Académic Française's Dictionaries run on Linux with the software "Philologic" (ATE metadata)
- The entire corpus is in French

OLAC Launch, LREC-02

FRANTEXT Catalog Metadata Format

<REF><SRF>M277</SRF><CON></CON><SRF>M278</SRF></REF>

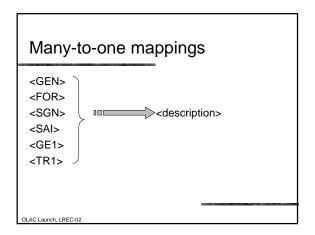
- <AUT><SAU><NOM>HUGO</NOM><PRE>Victor</PRE></SAU></AUT>
 <TIT>LA LEGENDE DES SIECLES</TIT>
- <DAT>1859</DAT>
- <TR1>non</TR1>
- <EDI>ED. P. BERRET, T.1 ET 2. PARIS : HACHETTE,1920.</EDI>
- <GEN><FOR><SGN>vers</SGN></FOR>
- <GE1><SGN>poésie</SGN></GE1></GEN>
- <SIE>19</SIE>
- <SAI>intégrale</SAI>
- <NBM>093885</NBM>
- <PUB>1</PUB>

OLAC Launch, LREC-02

Mapping ATILF to OLAC: First Experiment

- One-to-one mappings
- Many-to-one by collapsing to a single OLAC element
- OLAC refinements

OLAC Launch, LREC-02



OLAC mappings for FRANTEXT

<language code="fr"/>

<type.linguistic>corpus<type.linguistic>

<contributor refine="informant">ATILF</contributor>

OLAC Launch, LREC-02

Possible futures

- Enhance the ATILF metadata records
- Use new mappings: ATILF-TEI to OLAC
- Migrating the ATILF implementation from VIDA to Conventional by using scripting languages to describe response to harvest request verbs

OLAC Launch, LREC-02