PALM (Plateforme d’Analyse Linguistique Médiévale) : Morphology of Middle French Verbs with NooJ

Mourad Aouini, CLT, CNRS
Laure-Anne Caraty, Sorbonne Paris IV
Introduction
Palm ?
| Concordances | Articulate floribus moralibus debet erigi, uidelicet, quatuor virtutibus cardinalibus. | parte |
| In anteriori | ornari, scilicet, recordatione praeceptorum, circumspectione presentium, et prouidantia futurorum. Et ex | parte |
| sunt audacia in aggreiendo, patientia in sufferendo, et perseverantia in continuando. Et ex | parte |

<table>
<thead>
<tr>
<th>Annoter une forme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forme:</strong></td>
</tr>
<tr>
<td><strong>Lemme:</strong></td>
</tr>
<tr>
<td><strong>Langue:</strong></td>
</tr>
<tr>
<td><strong>Catégorie grammaticale:</strong></td>
</tr>
<tr>
<td><strong>Annotation(s) proposée(s):</strong></td>
</tr>
</tbody>
</table>
It contains:
- Base of texts written in French medium (XIV and XV centuries)
- Orientation of texts = politics (letters, charters, prince's mirrors, treaties, speeches etc.).

Integrated with NooJ formalism
Nooj?

Why Nooj in particular?

- Convenient as part of this app
- Description of the entire vocabulary of a language within a single dictionary
- Association with a lexical entry different orthographic variants
What can NooJ do for us?

• Teache the computer to relate each verb to a typical paradigm
• = a standard conjugation paradigm.

And how do we do that?

1- Write inflectional digital grammars
   = grammars of verbal inflection

2- Formalize the standard inflectional scheme of certain verbs

3- Take into account the different variations
Middle french what is it?

- state of the French language of the 14th and 15th centuries
- state of creation language subject to variation
  - Great phonetic changes
  - Fall of the declination of the old French
  - Simplification of variations of Old French verbal paradigms

1- for verbal radicals

\[ \text{afr} = \text{two categories: strong verbs with a variation of accent and homogeneous weak verbs = the variation of the accent causes form changes.} \]

ex: \( \text{amer} (=\text{aimer}) \) aim, aimes, aime, amons, amez, aiment

\( \text{morir} = \text{muir, muers, muert, morons, morez, muerent} \)

\( \text{parler} = \text{parol, paroles, parole, parlons, parlez, parolent} \)

2- for disinential morphemes (= endings)

\[ \Rightarrow \text{Neologism training from Latin} \]
The spelling: a new source of variation

1 the problem with abbreviations
→ Forget the meaning of the abbreviations of the old French
→ Add letters to these abbreviations
→ Result: redundancy!

ex: $x = \text{us} - \text{chevax}$ in old french
- $\text{chevaux}$ in middle french. $\text{chevaulx}$ sometimes!

2 latinizing spells
Insertion of consonants (not pronounced) in a word to recall its Latin origin
= Etymologizing spelling or etymological letter
- Result: doublets composed of one word with etymologizing spelling and another without etymologizing spelling
ex: redoubter / congnoistre
3 diacritical letter
Reading aid by modification of letter spelling
the i becomes y. See the rule of orthographia gallica :
« Quandoque i stat immediate ante vel post m, n vel u, potest mutari in y ut legibilior sit vel stare in sua natura. » (difficulty reading the legs !)

4 calligraphic letters (beautiful for writing)
Letters like Y more pleasing to the eye
Use without diacritical function
ex: in the verbal endings alternation / competition between i and y = oie against oye
Conclusion: shared complex language
  . simplify
  . spelling produces new variants
Example of text in Middle French
Christine de Pisan, *Les cent ballades d’amant et de dame*

Et Amours – Dieu le lui mire –
Quant de m’amour ay fait sire,
Sans desdire,
Tel que grant joye en recueil
Et a qui plaist sans desveil
Ce que je vueil.
Si me suis toute ordonnée
A l’amor, ne deffinee
Ne finee
N’iert ja l’amour, qui souffire
Me doit bien, car je me mire
Et remire
En sa beaute sans orgueil,
Et il fait, en tout accueil,
Ce que je vueil.
Prince, je suis sur le sueil
De joye quant voy a l’ueil
Ce que je vueil.
Our NooJ middle French dictionary

- Entries DMF (removal of unwanted forms)
- Integration of the entries of the Anglo-Norman Dictionary
- Graphic variants lexemes basic and common words
- For proper names: glossaries, indexes editions, encyclopedic information
- Variations and conjugated forms of verbs
- The number of dictionary entries: around 107,000 distinct forms (around 16.8% of anglo-normand forms)
Which verbs / How many verbs do we process

Limitation of the scope of work

• verbs of the first group.
• 50 recurring verbs in the corpus.
• Processing of **simple indicative tenses** (present, imperfect, future, conditional, simple past).
• Empirical approach.

Entrer, porter, jeter, achever, aymer/aimer, aider, otroier, laisser, envoyer, cuidier, aorer, apoiier, araisnier, mangier, ovrer, disner, plorer, rover, joer, demorer, esperer, etc.
But: great difficulty in processing so many verbs which are all from the 1st group!

1- Alignment of progressive paradigms over two centuries
2- Variation of the old French still present
3- Variation of bases still relevant for certain texts
4- Especially important variation in present tense indicative

<table>
<thead>
<tr>
<th>Number</th>
<th>Endings</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>-e <em>ou</em> rien</td>
</tr>
<tr>
<td>P2</td>
<td>-es</td>
</tr>
<tr>
<td>P3</td>
<td>-e <em>ou</em> –et <em>ou</em> rien</td>
</tr>
<tr>
<td>P4</td>
<td>-ons / –on</td>
</tr>
<tr>
<td>P5</td>
<td>-ez /-iez</td>
</tr>
<tr>
<td>P6</td>
<td>-ent/-ont</td>
</tr>
</tbody>
</table>
To summarize: huge variation if we take into account all the morphological parameters of the verbs that we are trying to reconcile with a computer construction

- Basic verbs: back1 pour P1 et P3 + Ø = type 1
- Basic verbs: Back 2 pour P1 / Back 1 pour P3 + Ø (=e) = type 2
- Two-base verbs: back 1 pour P3 (=e) et P5 + z (=ez) = type 3
- Verbs with three bases 2 back / 1back. No particular variation in person morphograms for frequent verbs. = type 4
- Two-base verbs: 2 back pour P1 et P3 + Ø = type 5
- Verbs with one / two bases (type –iier / -ier): 3 back pour P1 +Ø et P3 + -e, P5 en –iez = type 6
- Verbs with one / two bases (type –ier): 3 back pour P1 et P3 + -e, P5 en –iez = type 7
- Verb with two bases (type –ier): 3 back pour P1 + Ø et back 1 pour P5 + z (=iez) = type 8
Type 1

Porter

P1. port-e → B1 + suppr (r) + ajout Ø
P2. port-es → B1 + suppr (r) + ajout (-s)
P3. port-e → B1 + suppr. (r) + ajout Ø
P4. port-ons → B2 + suppr. (-er) + ajout (-ons)
P5. port-ez → B1 + suppr. (r) + ajout (-z)
P6. port-ent → B1 + suppr. (-r) + ajout (-nt)

Type 2

Torner

P1. torn → B2 + suppr. (-er) + ajout Ø
P2. torn-es → B1 + suppr. (r) + ajout (-s)
P3. torn-e → B1 + suppr. (-r) + ajout Ø
P4. torn-ons → B2 + suppr. (-er) + ajout (-ons)
P5. torn-ez → B1 + suppr. (-r) + ajout (-z)
P6. torn-ent → B1 suppr (-r) + ajout (-ent)

Type 3

Amer

P1. aim + Ø → B2 + ajout (Ø)
P2. aim-es → B1 + ajout (-s)
P3. aim-e → B1 + ajout Ø
P4. am-ons → B2 + suppr. (-er) + ajout (-ons)
P5. am-ez → B1 + suppr (-r) + ajout (-z)
P6. aim-ent → B1 + ajout (-nt)

Type 4

Preiser

P1. pris → B2 + Ø
P2. pris-es → B1 + ajout (-s)
P3. pris-e → B1 + suppr. (-r) + ajout Ø
P4. pris-ons → B2 + suppr. (-er) + ajout (-ons)
P5. pris-ez → B1 + suppr. (-r) + ajout (-z)
P6. pris-ent → B1 + ajout (-nt)

Type 5

Lever

P1.lief → B2 + Ø
P2.liev-es → B1 + ajout (-s)
P3.liev-e → B1 + suppr. (-r) + ajout Ø
P4.liev-ons → B2 + suppr. (-er) + ajout (-ons)
P5.liev-ez → B1 + suppr. (-r) + ajout (-z)
P6.liev-ent → B1 + ajout (-nt)

Type 6

Otroiier

P1.otroi → B3 + suppr. (-ier) + ajout Ø
P2.otroi-es → B3 + suppr. (-ier) + ajout (-es)
P3.otroi-e → B3 + suppr. (-ier) + ajout (-e)
P4.otroi-ons → B3 + suppr. (-ier) + ajout (-ons)
P5.otroi-ez → B3 + suppr. (-ier) + ajout (-ez)
P6.otroi-ent → B3 + suppr. (-ier) + ajout (-ent)

Type 7

Changier

P1.chang-e → B3 + Ø
P2.chang-es → B3 + suppr. (-ier) + ajout (-es)
P3.chang-e → B3 + suppr. (-ier) + ajout (-e)
P4.chang-ons → B3 + suppr. (-ier) + ajout (-ons)
P5.chang-iez → B3 + suppr. (-ier) + ajout (-iez)
P6.chang-ent → B3 + suppr. (-ier) + ajout (-ent)

Type 8

Cuidier

P1.cuit → B3 + Ø
P2.cuid-es → B3+ suppr. (-ier) + ajout (-es)
P3.cuid-e → B3 + suppr. (-ier) + ajout (-e)
P4.cuid-ons → B3 + suppr. (-ier) + ajout (-ons)
P5.cuid-iez → B1 + suppr. (-r) + ajout (-z)
P6.cuid-ent → B3 + suppr. (-ier) + ajout (-ent)
Highly complex system to establish grammars on NooJ!

What to do??

- Simplify the flex criteria with a focus on common rules
- Leaving aside the elements related only to the history of the language
- Take a purely formalistic point of view

- Search for common rules describing verbal inflection in Middle French
- Rules that can form other, simpler groups for writing numerical grammars
Identify two major morphological groups which make it possible to form two groups with a stable back number

<table>
<thead>
<tr>
<th>Verb in -er</th>
<th>Verb in -ier</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No longer take into account the number of bases</td>
<td></td>
</tr>
<tr>
<td>• <strong>Stick</strong> to a stable back count</td>
<td></td>
</tr>
<tr>
<td>• Choose a principle to select the number of backs:</td>
<td></td>
</tr>
<tr>
<td>• The greatest number of the paradigm</td>
<td></td>
</tr>
<tr>
<td>• Choose two standard verbs:</td>
<td></td>
</tr>
<tr>
<td><strong>entrer</strong> et <strong>changier</strong></td>
<td></td>
</tr>
</tbody>
</table>
Main graph « ENTREER »

French inflectional grammar consists of 6 graphs.
Indicative future
CONCLUSION

-Automatic processing of verbal inflection MF simplification / standardization of paradigms

-Impossible to standardize without forgetting HDL parameters

-Need for change of point of view for the medievalist
-NooJ = good current tool for MF treatment
- NooJ currently usable because most efficient technology
Bibliographie


Palm?

Platform was created in 2010

Online platform developed for:
- Development of digital texts and their sharing between researchers
- Analysis of texts in medieval Latin, English and Italian

Platform that allows:
- Automatic text annotations and
- Much easier manual corrections

Who is it aimed at?
- Historians, literary persons or philosophers who are not specialists in language

Goal?
- Use of computer tools for statistical and semantic analysis of late medieval text corpus
Verb in -er

- *Entrer:* largest back number: 2 at P4

*Entrr – ons suppr. er + -ons*

→ Consider the possible variations of the terminations at all times!
→ Check if the verbs match each time

Verb in -ier

*Changier:* greatest number of backs: 3 at P4

*Chang –ons suppr. ier + -ons*

Results:
- Uniform base
- Number of standardized backs for each group
- Variations saved only for completions