

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 ?



Plateforme d'analyse linguistique médiévale

Librairie Espace de travail Analyses Linguistiques Adm

Fréquences des formes

Forme	Fréquence
parte	4
regni	10
siue	4
corone	4
debet	6
Et	4
sic	6
corona	4
et	43
cuius	5
predicti	4
scilicet	4
que	8
iusticia	4
pro	4
ac	8
ad	6

Annotateur

... Ysaie . LXII . , Corona regni in manu Dei .
Et pro introductione materie textus illius annotauit ;
Cristiani in baptisate , in cuius signum unguentur et
signum gerunt tonsuram . Tertio , reges inuncti , in c

Annoter une forme

Forme: regni
Lemme: regnum
Langue: Latin
Catégorie grammaticale: Nom Commun
Annotation(s) proposée(s): Aucune

Valider

Annuler

Page

... Ysaie

Et pro

Cristia

signu

figura

gemm

precid

regen

seu al

prosp

In flor

preemin

cardinal

preterito

circumspe

et prouid

etiam gem

que sunt

audacia

in aggre

patientia

in suffer

et perse

uerantia

in contin

uando .

Et ex parte sinistra , debet poni temperantia , tribus gemmis ornata , ut restringat sensualitatem in uictu , refrenat loquelam in dictu , et uoluntatem suam siue desiderium in luxu .

In posteriori parte , poni debet iusticia , cuius tres sunt gemme , scilicet , ut fiat iusticia superioribus , equalibus , et inferioribus ; dixit que ulterius , quod si corona moralis alicuius regni sic disponatur , concludi potest , quod prius assumitur corona regni in manu Dei : quibus premissis sic annotatis , et per auctoritates pagine diuine , alias que notabilitates quamplurimas egregie uallatis et roboratis ; prefatus cancellarius , tres causas summonitionis parliamenti predicti notabiliter exposuit et publicauit .

Prima scilicet causa , pro iusticia et pace inter subditos et ligeos regis regni predicti enutriendis , manutenendis et supportandis ; ac cuilibet eorumdem equanimiter ministrandis . Secunda , ad cognoscendum per auisamentum ac sanas discretiones dominorum et communitatis predictorum , uias et media , qualiter et quo modo mercandise de crescentia regni

ones homini coronantur . Primo scilicet ,

in sacris ordinibus constituti , in cuius

uro et gemmis ornata , in cuius corone

otatur , et in floribus corone erectis et

nam sicut aurum est metallum maxime

abilis in se , et in fidelitate penes suum

in flexible et ductile , ad formam corone

is et ductilis , ad regis honorem et regni

Palm ?

Concordances

- | | | |
|--|-------|---|
| 1. , que quatuor floribus moralibus debet erigi , uidelicet , quatuor uirtutibus cardinalibus
. In anteriori | parte | corone, debet poni prudentia, que tribus gemmis debet ornari, |
| 2. ornari , scilicet , recordatione preteritorum , circumspectione presentium , et prouidentia futurorum . Et ex | parte | dextera, debet erigi fortitudo, tribus etiam gemmis ornata, que |
| 3. sunt audacia in aggrediendo , patientia in sufferendo , et perseuerantia in continuando . Et ex | parte | sinistra, debet poni temperantia, tribus gemmis ornata, ut restringat |
| 4. posteriori | parte | , poni debet iusticia, cuius tres sunt gemme, scilicet, |

ordatione preteritorum , circumspectione presentium , et um . Et ex	parte	dextera, debet erigi fortitudo, tribus etiam gemmis ornata, que
rediendo , patientia in sufferendo , et perseuerantia in	parte	sinistra, debet poni temperantia, tribus gemmis ornata, ut restringat
	parte	, poni debet iusticia, cuius tres sunt gemme, scilicet,

Annoter une forme

Forme: parte

Lemme: pars

Langue: Latin

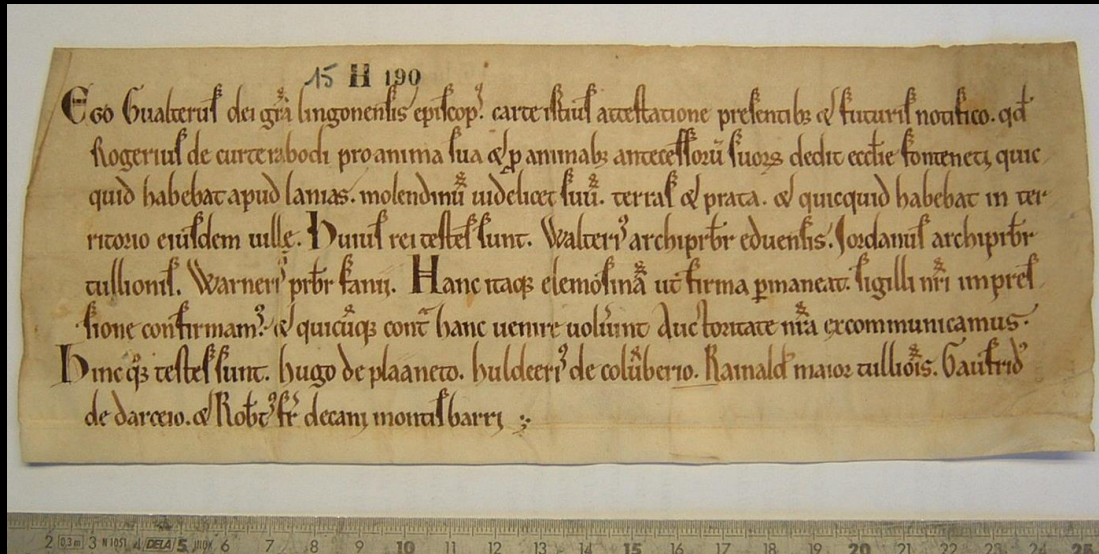
Catégorie_grammaticale: Préposition

Annotation(s)_proposée(s):
Préposition
Conjonction de subordination
Conjonction de coordination
Interjection
Nom Propore
Pronom
Nom Commun
Nombre
Verbe
Adjective
Adverbe

Valider

Corpus ?

- Imported into the application
- Those of the MEDITEXT internal library



Integrated with NooJ formalism

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.).

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?

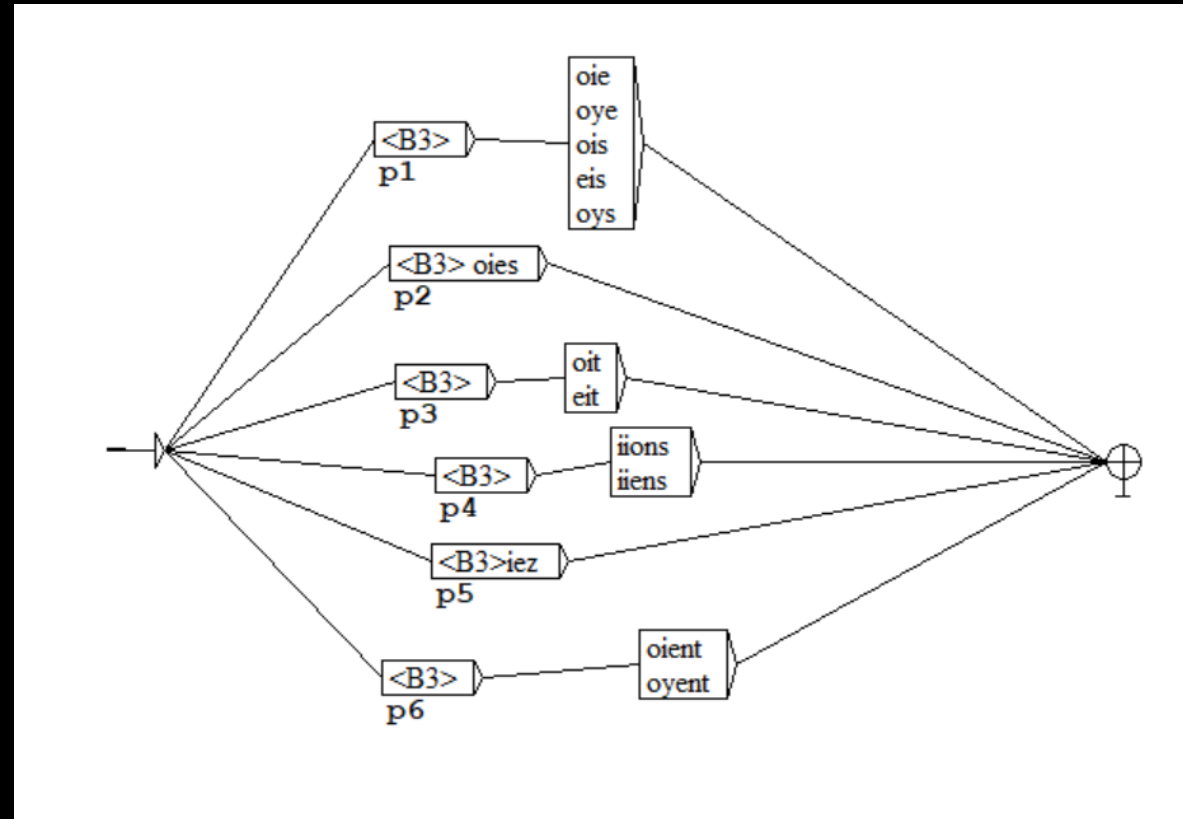
- Teache the computer to relate each verb to a typical paradigm
- = a standard conjugation paradigm.

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

afr =two categories: strong verbs with a variation of accent and
homogeneous weak verbs = the variation

of the accent causes form changes.

ex: amer (=aimer) aim, aimes, aime, amons, amez, aiment

morir = muir, muers, muert, morons, morez, muerent

parler = parol, paroles, parole, parlons, parlez, parolent

2- for disinential morphemes (= endings)

→ 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 = us - *chevax* in old french

- *chevaux* in middle french . *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*

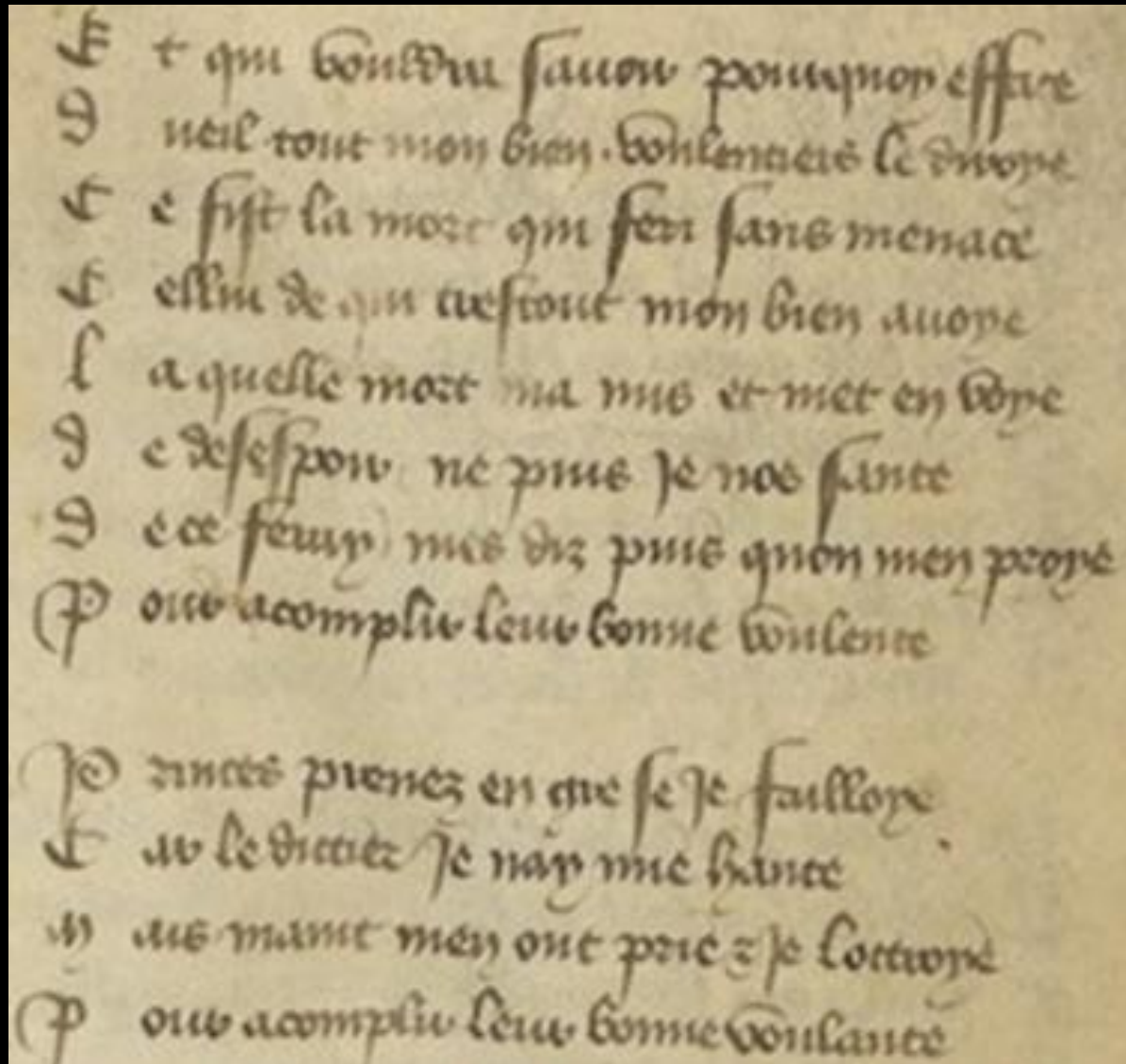
Conclusion: shared complex language

. 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 ordonnee
A l'**amer**, ne deffinee
Ne finee
N'**iert** ja l'amour, qui souffire
Me doit bien, car je me mire
Et remire
En sa beauté 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.



Et qui voudra sauoir pourquoy effire
D'ueil tout mon bien, contentiers le d'uoire
Et fist la mort qui fer sans menace
L'ellui de qui trestout mon bien auoie
La quelle mort ma mis et met en uoie
D'e desespoir ne puis je nos fante
D'ee feuy mes diz puis qu'on men peone
Pour acomplir leur bonne uolente

Princes prenez en que se je faulloye
Et au le dictiez je nay mie hante
Mais maint men ont priez je loctioye
Pour acomplir leur bonne uolante

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, acheter, aimer/aimer, aider, otroier, laisser, envoyer, cuidier, aorer, apoier, 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

Theoretical table of endings from present indicative to first group

P1: -e *ou* rien

P2: -es

P3: -e *ou* -et *ou* rien

P4: -ons / -on

P5: -ez /-iez

P6: -ent/-ont

Result

- Otroier
- P1. otroi .+ ajout Ø
- P2. otroi-es
- P3. otroi-e
- P4. otroi-ons
- P5. otroi-ez
- P6.otroi-ent

- Laissier
-
- P1. laiss-e
- P2. laiss-es
- P3.laiss-e
- P4. laiss-ons
- P5. laiss-iez
- P6. laiss-ent

- Changier;
-
- P1. chang-e
- P2.chang-es
- P3.chang-e
- P4.chang-eons
- P5.chang-iez
- P6.chang-ent

- Preiser: B1 preis- ; B2 pris-
-
- P1. pris + Ø
- P2.pris-es
- P3.pris + Ø
- P4.preis-ons
- P5.preis-ez
- P6.pris-ent

- Parler: B1 parl- ; B2 parol-
-
- P1. parol + Ø
- P2.parol-es
- P3.parol-e
- P4.parl-ons)
- P5.parl-ez
- P6.parol-ent

•Chanter

- P1. chant + Ø
- P2.chant-es
- P3.chant-e
- P4.chant-ons
- P5.chant-ez
- P6.chant-ont

- Mangier : B1 manj- ; B2 manju-
-
- P1. manju + Ø
- P2.manju-es
- P3.manju-e
- P4.manj-ons
- P5. manj-iez

- Comparer: B1 compar- ; B2 comper-
-
- P1. comper-e
- P2.comper-es
- P3.comper-e
- P4.compar-ons
- P5.compar-ez
- P6.comper-ent

- Ovrer: B1 ovr-; B2 uevr-
-
- P1.uevr-e
- P2.uevr-es
- P3.uevr-e
- P4.ovr-ons
- P5.ovr-ez
- P6.uevr-ent

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 + \emptyset = type 1
- Basic verbs : Back 2 pour P1 / Back 1 pour P3 + \emptyset (= -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 + \emptyset = type 5
- Verbs with one / two bases (type -iier / -ier): 3 back pour P1 + \emptyset 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 + \emptyset et back 1 pour P5 + z (= iez) = type 8

<p>Type 1 •Porter</p> <p>•P1. port-e → B1 + suppr (-r) + ajout Ø</p> <p>•P2. port-es → B1 + suppr (-r) + ajout (-s)</p> <p>•P3. port-e → B1 + suppr. (-r) + ajout Ø</p> <p>•P4. port-ons → B2 + suppr. (-er) + ajout (-ons)</p> <p>•P5. port-ez → B1 + suppr. (-r) + ajout (-z)</p> <p>•P6. port-ent → B1 + suppr. (-r) + ajout (-nt)</p>	<p>Type 2 •Torner</p> <p>•P1. torn → B2 + suppr. (-er) + ajout Ø</p> <p>•P2. torn-es → B1 + suppr. (-r) + ajout (-s)</p> <p>•P3. torn-e → B1 + suppr. (-r) + ajout Ø</p> <p>•P4.torn-ons → B2 + suppr. (-er) + ajout (-ons)</p> <p>•P5.torn-ez → B1 + suppr. (-r) + ajout (-z)</p> <p>•P6.torn-ent → B1 suppr (-r) + ajout (-ent)</p>	<p>Type 3</p> <p>•Amer</p> <p>•P1. aim + Ø → B2+ ajout (Ø)</p> <p>•P2. aim-es → B1 + ajout (-s)</p> <p>•P3. aim-e → B1 + ajout Ø</p> <p>•P4. am-ons → B2 + suppr (-er) + ajout (-ons)</p> <p>•P5. am-ez → B1 + suppr (-r) + ajout (-z)</p> <p>•P6. aim-ent → B1 + ajout (-nt)</p>	<p>Type 8 •Cuidier</p> <p>•P1. cuit → B3 + Ø</p> <p>•P2. cuid-es → B3+ suppr. (-ier) + ajout (-es)</p> <p>•P3.cuid-e → B3 + suppr. (-ier) + ajout (-e)</p> <p>•P4.cuid-ons → B3 + suppr. (-ier) + ajout (-ons)</p> <p>•P5.cuid-iez → B1 + suppr. (-r) + ajout (-z)</p> <p>•P6.cuid-ent → B3 + suppr. (-ier) + ajout (-ent)</p>
<p>Type 4 •Lever</p> <p>•P1.lief → B2 + Ø</p> <p>•P2.liev-es → B1 + ajout (-s)</p> <p>•P3.liev-e → B1 +suppr -r ajout Ø</p> <p>•P4.lev-ons → B2 + suppr (-er) + ajout (-ons)</p> <p>•P5.lev-ez → B1 + suppr (-er) + ajout (-z)</p> <p>•P6.liev-ent → B1 + ajout (-nt)</p>	<p>Type 5</p> <p>•Preiser</p> <p>•P1. pris → B2 + Ø</p> <p>•P2.pris-es → B1 + ajout (-s)</p> <p>•P3.pris → B2 + Ø</p> <p>•P4.preis B2 + ons</p> <p>•P5.preis B1 + -z</p> <p>•P6.pris-ent → B1+ ajout (-nt)</p>	<p>Type 6</p> <p>•Otroier</p> <p>•P1. otroi → B3 + suppr. (-ier) + ajout Ø</p> <p>•P2. otroi-es → B3 + suppr. (-ier) + ajout (-es)</p> <p>•P3. otroi-e → B3 + suppr. (-ier) + ajout (-e)</p> <p>•P4. otroi-ons → B3 + suppr. (-ier) + ajout (-ons)</p> <p>•P5. otroi-ez → B3 + suppr. (-ier) + ajout (-ez)</p> <p>•P6.otroi-ent → B3 + suppr. (-ier) + ajout (-ent)</p>	<p>Type 7 •Changier</p> <p>•P1. chang-e → B3 + suppr. (-ier) + ajout (-e)</p> <p>•P2.chang-es → B3 + suppr. (-ier) + ajout (-es)</p> <p>•P3.chang-e → B3 + suppr. (ier) + ajout (-e)</p> <p>•P4.chang-eons → B3 + suppr. (-ier) +e+ ajout (-ons)</p> <p>•P5.chang-iez → B3+ suppr. (-ier)+ ajout (-iez)</p> <p>•P6.chang-ent → B3 + suppr. (-ier) + ajout (-ent)</p>

**Highly complex system
to establish grammars
on NooJ !**

What to do ??

- **Search for common rules describing verbal inflection in Middle French**
- **Rules that can form other, simpler groups for writing numerical grammars**

- 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**

Identify two major morphological groups which make it possible to form two groups with a stable back number

Verb in -er

Verb in -ier

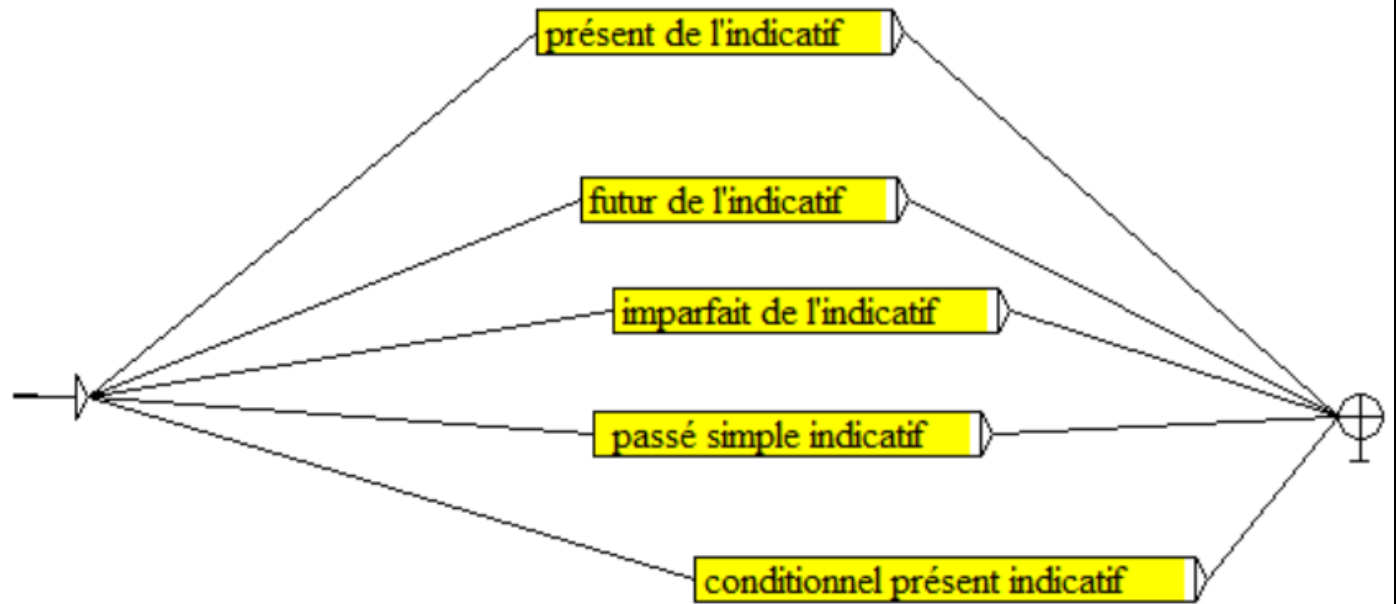
- No longer **take** into account the number of bases
- **Stick** to a stable back count
- Choose a principle to select the number of backs:
- The greatest number of the paradigm
- Choose two standard verbs :
entrer et changer

Main graph « ENTRER »

Main

- ... passé simple indicatif
- ... conditionnel présent indicatif
- ... futur de l'indicatif
- ... imparfait de l'indicatif
- ... présent de l'indicatif

French inflectional grammar consists of 6 graphs.

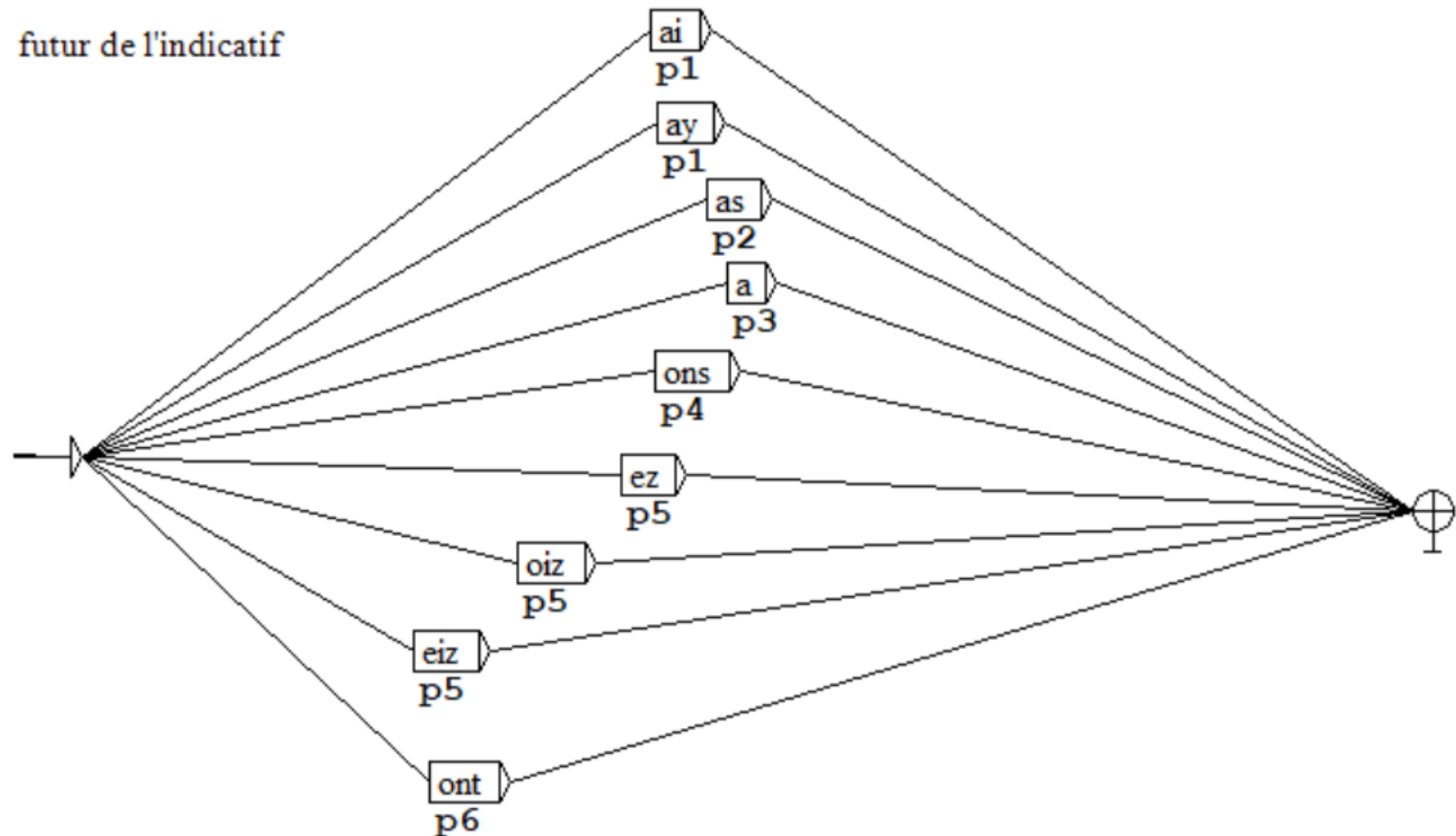


Indicative future

Main

- passé simple indicatif
- conditionnel présent indicatif
- futur de l'indicatif
- imparfait de l'indicatif
- présent de l'indicatif

French inflectional grammar consists of 6 graphs.



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

(Silberztein, 2015) SILBERZTEIN, Max. *La formalisation des langues: l'approche NooJ*. ISTE ed., 2015.

(Aouini, 2014) AOUMINI, Mourad. A NooJ module for named entity recognition in middle french. *Formalising Natural Languages with Nooj 2014*, 2015, p. 99.

(Aouini, 2018) AOUMINI, Mourad. Approche multi-niveaux pour l'analyse des données textuelles non-standardisées: corpus de textes en moyen français. 2018. Thèse de doctorat. Franche-comte.

(Mesfar, 2008) MESFAR, Slim. *Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard*. 2008. These de doctorat. Besancon.

(Mikolov et al., 2013) MIKOLOV, Tomas, SUTSKEVER, Ilya, CHEN, Kai, *et al.* Distributed representations of words and phrases and their compositionality. In : *Advances in neural information processing systems*. 2013. p. 3111-3119.

(Hetsevich et al., 20) Hetsevich, Yuras, Kryvaltsevich, Alena, Kazlouskaja, Nastassia. Sentiment Analysis Algorithms for the Belarusian NooJ Module in Touristic Sphere.

(Bisazza, 2009) Bisazza, Arianna. Designing a NooJ Module for Turkish Inflectional Analysis: an Example of Highly Productive Morphology. In : NooJ Conference. 2009.

(Mesfar, 2011) Mesfar S. 2011. Towards a Cascade of Morpho-syntactic Tools for Arabic Natural Language Processing. Proceedings of CICLing 2010, Romania: 150-162

Bogacki, 2012) Bogacki K. and Gwiazdecka E. 2012. Derivational Structure of Polish Verbs and the Expansion of the Dictionary. Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2011 International Conference (Dubrovnik, Croatia). Edited by Kristina Vučković, Božo Bekavac and Max Silberztein. Cambridge Scholars Publishing, Newcastle., UK: 50-62

(Marchello-Nizia, 2005) Marchello-Nizia C.,2005, *La langue française aux XIVe et XVe siècles*. Armand Colin.

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

Entr – 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**