



Utilité de la linguistique de corpus et de la statistique. Exemples d'application à des textes français et japonais

Yuji KAWAGUCHI

Tokyo University of Foreign Studies, Japan

Sommaire

1. Cadre general de recherches
2. Analyse de la micro-variation graphique dans un manuscrit
3. Graphies médiévales et changements phonétiques en cours (1230-1300)
4. Témoignages du français parlé au XVIIIe siècle
5. Sociolinguistique et pragmatique du japonais au début XIXe siècle.

Conclusion



I. Cadre général de recherches

Situation actuelle

- Collaboration entre Informatique et Linguistique
 - Downsizing mécanique
- Changement de l'approche linguistique
 - Linguistique de corpus
 - Analyse qualitative et quantitative

The 21st century COE Program 'Usage-Based Linguistic Informatics' (2002-2006)

<http://www.coelang.tufs.ac.jp/english/index.html>

Usage-Based Linguistic Informatics

The aim of this project is to innovate foreign language education by developing superior foreign language educational material and transmitting it through the Internet.

(UBLL)

Thus, an overall integration of theoretical and applied linguistics will be realized through the theoretical analysis of linguistic usages of different languages.

- ① Collaboration entre informatique et linguistique
- ② Importance du corpus de la langue réelle
- ③ Mise en relation entre linguistique et didactiques de langue

Analyse linguistique basée sur le corpus

- synchronique
constitution de corpus

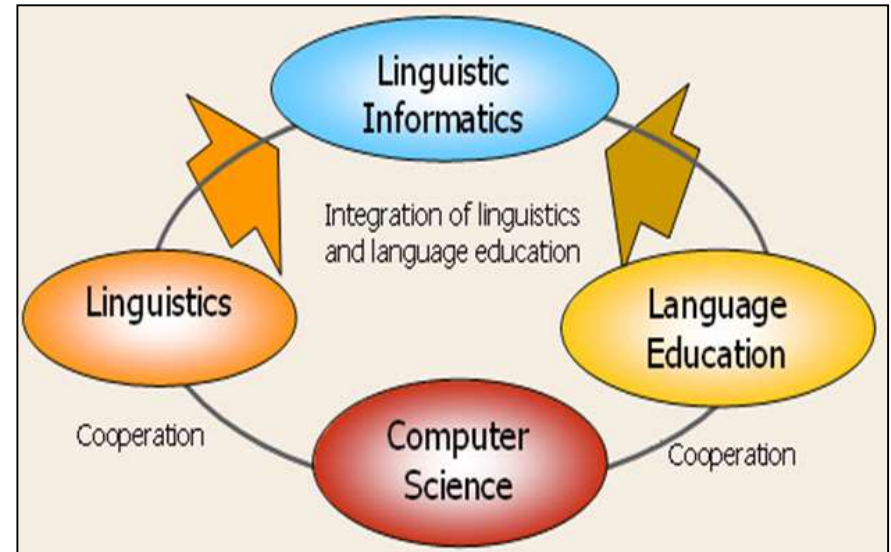
oraux de langues :

- ❑ français
- ❑ français canadien
(de locuteurs bilingues)
- ❑ malais
- ❑ espagnol
- ❑ russe
- ❑ turc
- ❑ japonais (de locuteurs natifs)
- ❑ japonais (d'apprenants)

- diachronique

quelques constitution de corpus

- ❑ français médiéval etc.
- ❑ russe médiéval
- ❑ Atlas Linguistique du français
- ❑ Atlas Linguistique du japonais



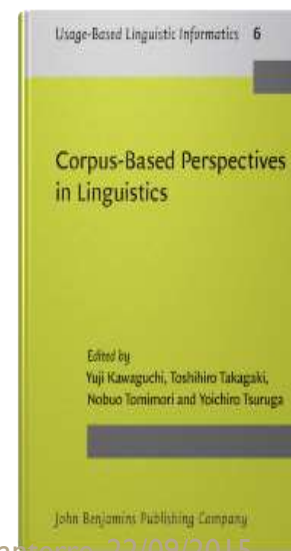
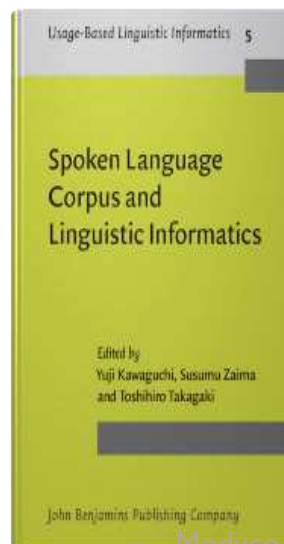
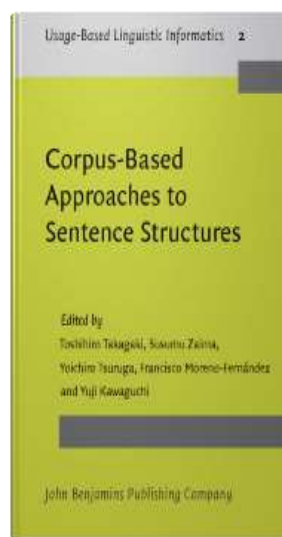
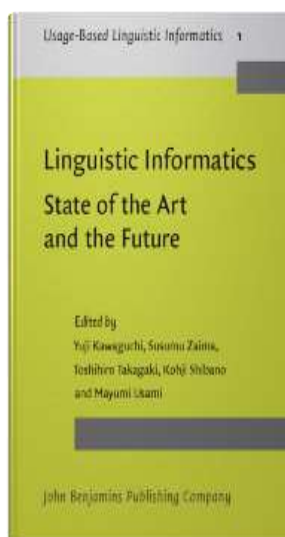
- didactiques de langue

The screenshot shows the 'TUFS Language Modules' website. The main heading is 'Français' with the Japanese text 'フランス語' next to it. Below the heading, there is a section titled 'Le monde francophone et sa population' with text about the French language's global presence. Another section titled 'Le français utilisé dans ce matériel pédagogique' describes the language used in the materials. On the left side, there is a sidebar with a 'Module de conversation' section and a list of language modules including 'Français', 'Français (Québec)', 'Français (Suisse)', 'Français (Inde)', and 'Module japonais'. There are also logos for 'Recycling University and Foreign Studies' and 'UNIVERSITÉ DE NANTERRE'.

TUFS Modules de langue

Series : Usage-Based Linguistic Informatics (UBLI), Amsterdam/Philadelphia, John Benjamins

6. *Corpus-Based Perspectives in Linguistics*, 2007. vi, 442 pp.
5. *Spoken Language Corpus and Linguistic Informatics*, 2006. vi, 434 pp.
4. *Readings in Second Language Pedagogy and Second Language Acquisition: In Japanese Context*, 2006. vi, 274 pp.
3. *Prosody and Syntax: Cross-linguistic perspectives*, 2006. vi, 384 pp.
2. *Corpus-Based Approaches to Sentence Structures*, 2005. vi, 317 pp.
1. *Linguistic Informatics – State of the Art and the Future: The first international conference*, 2005. viii, 363 pp.



Global COE Program 'Corpus Based Linguistics and Language Education' (2006-2012) <http://cblle.tufs.ac.jp/index.php>

Japanese

Global COE Program, Tokyo University of Foreign Studies
Corpus-based Linguistics and Language Education



Menu:

- Objectives
- Field Linguistics
- Corpus linguistics
- Linguistics Informatics
- Conferences
- Nurturing researchers
- Publications
- Core Members
- Appointed Professor

English Top

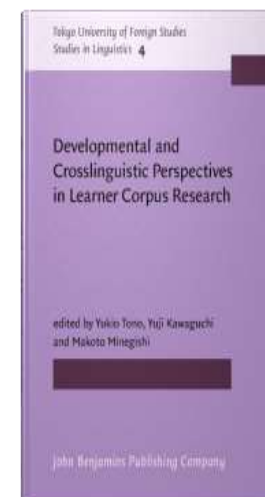
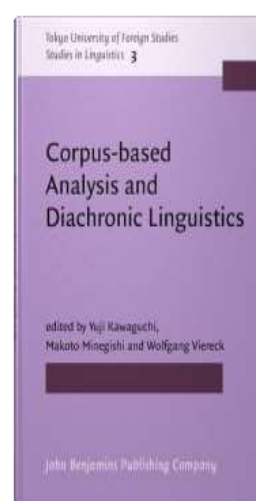
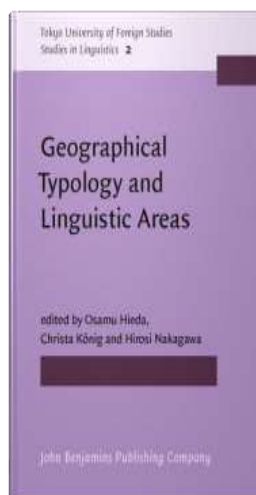
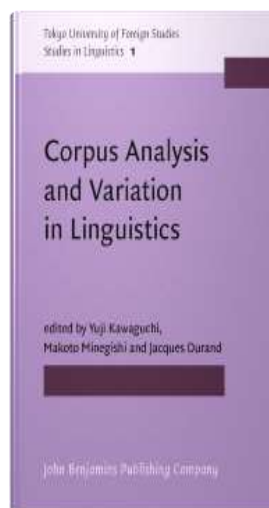
The objective of this Global COE Program, 'Corpus-based Linguistics and Language Education' (CblLE), is to foster internationally competitive researchers in subfields of linguistics that emphasize the empirical (particularly corpus-based) study of language structure and language education.

What's new

- 2011/06/29 **Working Papers in Corpus-based Linguistics and Language Education 7**
- 2011/04/20 **Handbooks in Corpus-based Linguistics and Language Education VI**
- 2011/03/29 **Working Papers in Corpus-based Linguistics and Language Education 6**
- 2010/10/15 **Interphonologie du Français Contemporain (IPFC)**
- 2010/10/13 **Handbooks in Corpus-based Linguistics and Language Education V**
- 2010/10/01 **French Conversation Module (TUFS LANGUAGE MODULES)**
- 2010/06/14 **Working Papers in Corpus-based Linguistics and Language Education 5**

Series :TUFS Studies in Linguistics (TUFS SL), Amsterdam/Philadelphia, John Benjamins

4. *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*, 2012. vi, 361 pp.
3. *Corpus-based Analysis and Diachronic Linguistics*, 2011. vi, 293 pp.
2. *Geographical Typology and Linguistic Areas: With special reference to Africa*, 2011. vi, 321 pp.
1. *Corpus Analysis and Variation in Linguistics*, 2009. vi, 399 pp.





2. Analyse de la micro-variation graphique dans un manuscrit



Corpus:

Treize Miracles de Notre-Dame (TM)

Ms. B.N. fr. 2094, Ottawa, Éditions de
l'Université d'Ottawa, 1981,
Pierre Kunstmann

3 manuscrits

A: Paris, Bibliothèque Nationale, français, 2094

B: Cambridge, Fitzwilliam Museum, McClean, 178

C: Carpentras, Bibliothèque Inguimbertaine, 106

Constitution du corpus

Miracle I

.....

[56] En l'ostel n'ot celi ne cele

[57] Qui ne seüst de verité

[58] Que la roïne de pité,

[59] La mere Deu, la debonere,

[60] Vout au pseudome ce don fere

[61] Por lui sooler et repestre.

[62] Cil lo conterent a lor prestre

[63] Qui virent lo miracle apert.

[64] Je di por voir que son tans pert

[65] [fol.151 v^a] Qui ne se porchace et acointe

[66] De ceste dame noble et cointe,

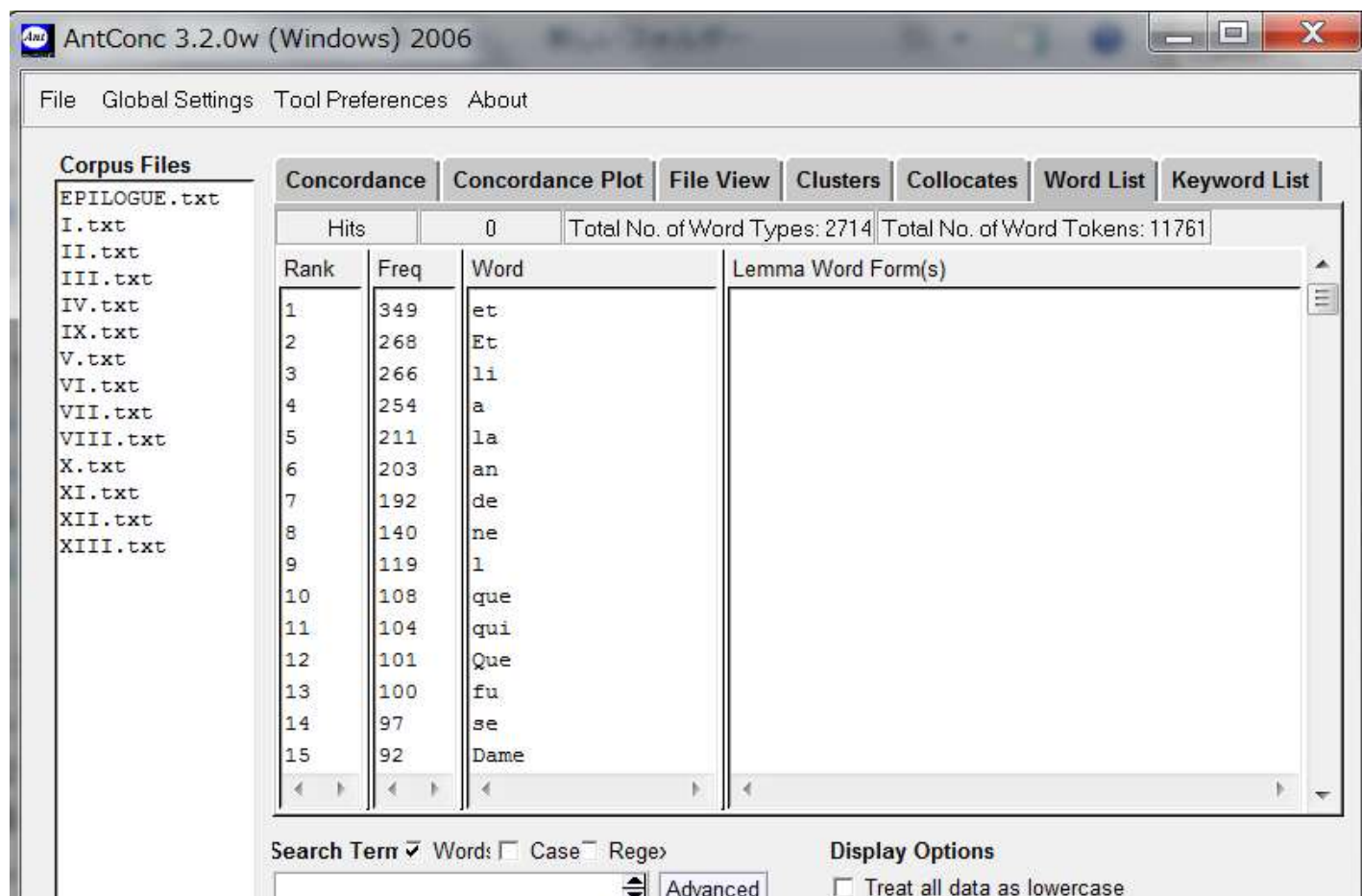
.....



logiciel Reconnaissance
Optique de Caractères
e Typist pour 58 langues

Concordancier *AntConc* et analyse fréquentielle

AntConc : <http://www.laurenceanthony.net/software.html>



AntConc 3.2.0w (Windows) 2006

File Global Settings Tool Preferences About

Corpus Files

- EPILOGUE.txt
- I.txt
- II.txt
- III.txt
- IV.txt
- IX.txt
- V.txt
- VI.txt
- VII.txt
- VIII.txt
- X.txt
- XI.txt
- XII.txt
- XIII.txt

Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Hits: 0 Total No. of Word Types: 2714 Total No. of Word Tokens: 11761

Rank	Freq	Word	Lemma Word Form(s)
1	349	et	
2	268	Et	
3	266	li	
4	254	a	
5	211	la	
6	203	an	
7	192	de	
8	140	ne	
9	119	l	
10	108	que	
11	104	qui	
12	101	Que	
13	100	fu	
14	97	se	
15	92	Dame	

Search Term ☒ Words ☐ Case ☐ Regexp ☐ Display Options

☐ Treat all data as lowercase

Advanced

Concordance

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

- EPILOGUE.txt
- I.txt
- II.txt
- III.txt
- IV.txt
- IX.txt
- V.txt
- VI.txt
- VII.txt
- VIII.txt
- X.txt
- XI.txt
- XII.txt
- XIII.txt

Concordance Concordance PI File Vie Clusters/N-Grar Collocat Word Li Keyword Li

Concordance Hits 6

Hit	KWIC	File
1	stoient bon crestien; [69] Avesques an ert sainz Bas	XII.txt
2	an sa maison [76] Vivoit l' avesques , li bons hon; [77	XII.txt
3	uce dame anorer; [145] Li avesques devant s'esmut	XII.txt
4	nalement. [154] Lors vit l' avesques visaumant [155	XII.txt
5	res : [180] « Ma dame, cit avesques a nostres, [181]	XII.txt
6	grant melodie. [205] Et li avesques s'esvoilla; [206]	XII.txt

avesques “évêque”

Search Term ☒ Words ☐ Case ☐ Regex

Search Win

avesques Advanced 50

Pointer les forms dans le corpus

The screenshot shows the Concordance software interface. The menu bar includes File, Global Settings, Tool Preferences, and Help. The 'Corpus Files' list on the left contains EPILOGUE.txt, I.txt, II.txt, III.txt, IV.txt, IX.txt, V.txt, VI.txt, VII.txt, VIII.txt, X.txt, XI.txt, XII.txt, and XIII.txt. The main window has tabs for Concordance, Concordance Plot, File View, Clusters/N-Gram, Collocate, Word List, and Keyword List. The 'Concordance Hits' section shows 6 hits and 1 total plot. The 'HIT FILE: 13' and 'FILE: XII.txt' are displayed. A table with 5 columns and 1 row is shown. To the right of the table, it says 'No. of Hits = 6' and 'File Length (in chars) = 9715'. The search results display the text 'avesques trouvé six fois dans un seul Miracle XII'. At the bottom, the 'Search Term' is 'avesques', with checkboxes for 'Words' (checked), 'Case', and 'Regex'. The 'Plot Zoom' button is also visible.

File Global Settings Tool Preferences Help

Corpus Files

- EPILOGUE.txt
- I.txt
- II.txt
- III.txt
- IV.txt
- IX.txt
- V.txt
- VI.txt
- VII.txt
- VIII.txt
- X.txt
- XI.txt
- XII.txt
- XIII.txt

Concordance Hits 6 **Total Plots** 1

HIT FILE: 13 FILE: XII.txt

--	--	--	--	--

No. of Hits = 6
File Length (in chars) = 9715

avesques trouvé six fois dans un
seul Miracle XII

Total No. **Search Term** ☒ Words ☐ Case ☐ Regex **Plot Zoom**

Taille du corpus

B.N. fr. 2094, considéré généralement
comme bourguignon

texte	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	Ep
n.t.m.	500	973	1045	729	523	871	612	1306	1405	710	613	1597	592	212

n.t.m. = nombre total de mots
Taille de 13 Miracles et un Epilogue

Questions de recherche

1. Peut-on considérer *TM* comme un texte linguistiquement homogène ?
2. Peut-on dire que *TM* sont les textes bourguignons ?

Rimes et langue d'auteur

Confusion entre *-ai/-e-* et *-a-*

saiges/passesges III[49-50]

sauvaige/boischage IX[21-22]

mariaige/voiage IX[75-76]

linage/saige X[26-27]

saiges/volaiges III[7-8]

saige/linaige III[73-74]

usaige/saige VIII[7-8]

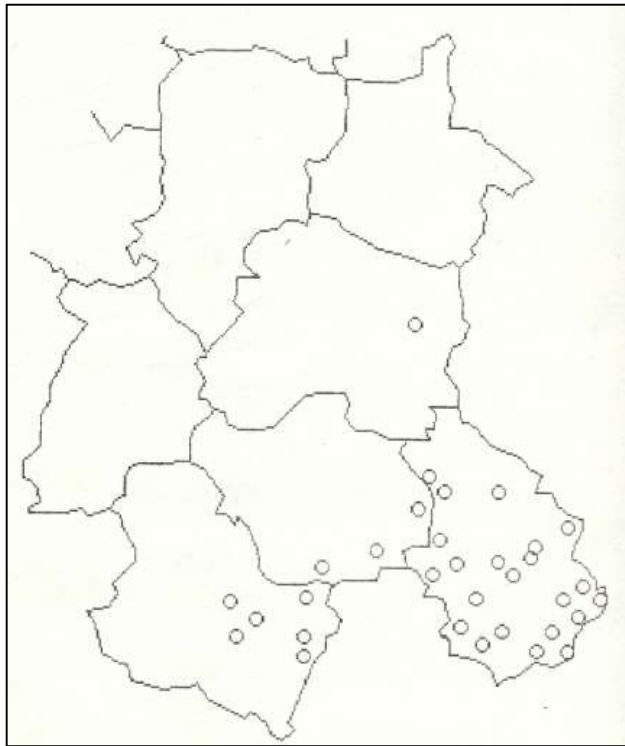
image/visaige IX[187-188]

mais *ymage/usage* IV[49-50]

Le chiffre roman I à XIII signale treize miracles de *TM* et *Ep* l'épilogue.

Le chiffre arabe représente le numéro du vers.

fromage > fromaige



légende

- [fròmèʒ] (1) [frò:mæ:ʒ] (1)
- [fròmèʒ] (4) [frœmèʒ] (13)
- [fròmèʒ] (4) [frœ:mæ:ʒ] (1)
- [frumèʒ] (8)

*Atals Linguistique et Ethnographique de
Champagne et la Brie (ALCB)*

n° 948 (LE) FROMAGE



légende

- [fròmèʒ] ou [frumeʒ]

*Atals Linguistique et Ethnographique de
Bourgogne (ALB)*

n° 1104 le fromage

Autres graphies régionales

Graphie *-auble* pour *-able*

parmenauble/tauble XIII[93-94]

mais *estable/estable* XII[233-234]

Absence de *-d-* transitoire

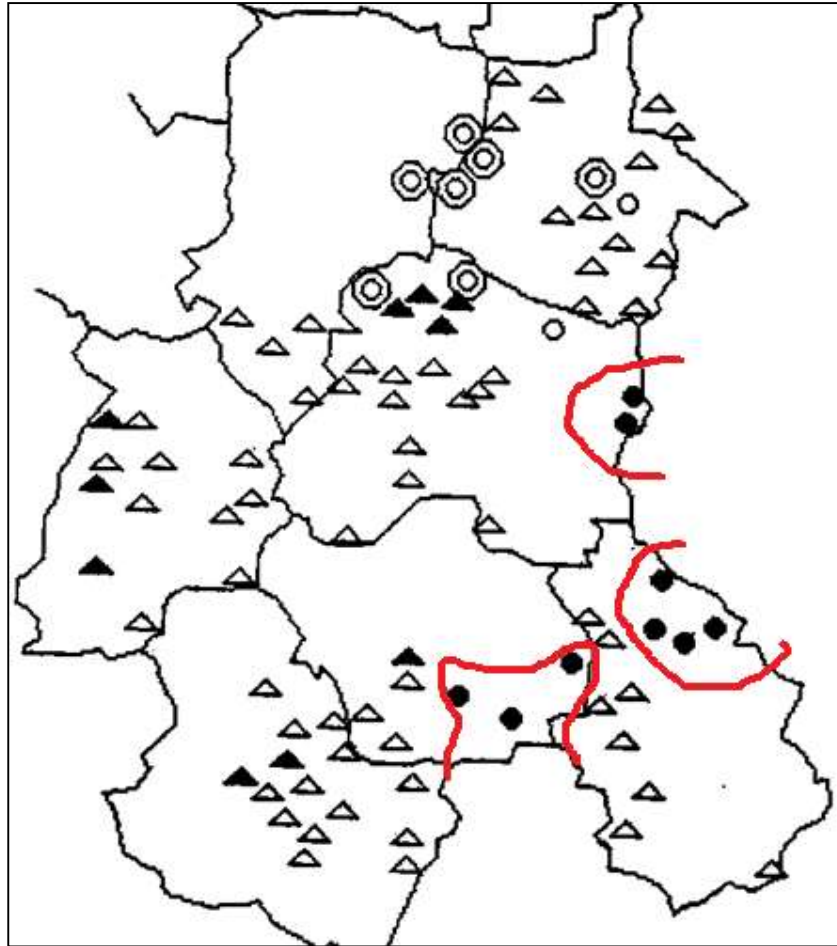
vanra/auras XI[89-90] et hors de rime *manra* XII[103]

mais *revendra* IX[89], *revendras* XI[85],

voudra IV[83], VI[29], *vendrons* XIII[89],

voudras IV[78], *Voudrons* VI[43]

-able > -auble



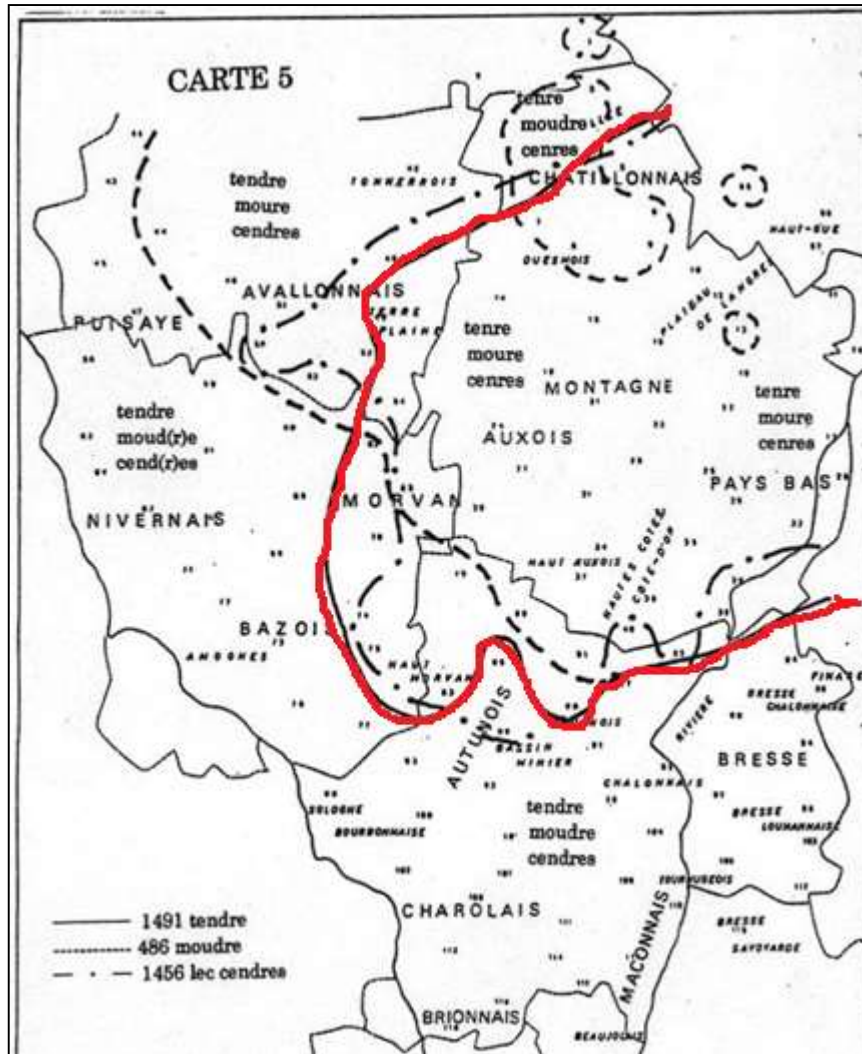
légende

- [èró:l] (4) [éró:l] (4) [érò:l] (1)
- [èró:b] (2)
- ◎ [èrø:l] (1) [érœl] (1) [érrø:l] (5)
- △ [èrá:b] (6) [èráb] (10) [érá:b] (43) [éràb] (2) [érá:b] (1)
- ▲ [éràbl] (3) [érábl] (5) [érá:bl] (2)

Atals Linguistique et Ethno-graphique de Champagne et la Brie (ALCB)

n° 558 (L')ERABLE

Absence de -d- transitoire



tenre = *tendre* < lat. *ten(e)re*
moure = *moudre* < lat. *mol(e)re*
cenres = *cendres* < lat. *cin(e)re*

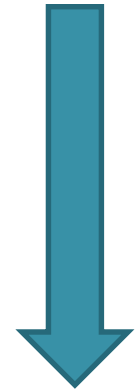
14 traits graphiques extraits

1. Confusion entre ai, e et a: aigniaus, aillons, amessent, ansarreras, avangiles, avesque «évêque», balances, j'a «j'ai», jama, je «ja», jema, lermes, lai «là», memales, senez, trabuchier, trare;
2. –aige «-age»: domaige, imaige, imayges, linaiges, mariaige, masaiges, mesaige, passeges, saige, saiges, sauvaige, usaige, visaige, volaigesm ymaige;
3. –arent «-erent»: alarent, donarent, lassarent, reconfortarent, s'acordarent;
4. –auble «-able»: deaubles, parmenauble, tauble;
5. –ierre «-erre»: pechierres;
6. –nr- «-ndr-»: manra, vanras;
7. chute de –r-: ariés, cetans, divés, jurés, martire, martis, pallerent, parlés «parlers», priox, tite, toz «torz», ves, vois «voirs»;
8. chute de –s-: c'et «c'est», cit «cist», m'et, s'et, tritece;
9. –omes: devonmes, fusomes, savommes;
10. arme;
11. yglise;
12. virge, virges;
13. boiche, roige;
14. consoil, consoilla, consoilliez, mervuille, oroilles.

Distribution de Bernoulli

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
I	0	0	0	0	0	0	0	0	0	0	0	1	0	0
II	2	1	0	0	1	0	1	1	0	1	0	0	0	3
III	0	8	0	0	1	0	0	0	1	1	1	0	1	2
IV	3	0	0	0	0	0	0	0	0	1	0	0	0	1
V	0	1	0	0	0	0	1	0	0	0	0	0	0	1
VI	0	1	1	0	0	0	2	0	0	0	0	0	4	2
VII	1	0	0	0	0	0	2	0	1	3	0	1	0	0
VIII	3	2	3	1	0	0	6	2	0	1	2	2	2	0
IX	8	4	1	1	0	0	1	3	0	3	1	2	0	0
X	1	2	0	0	0	0	2	2	0	1	0	1	0	2
XI	2	0	1	2	0	1	2	4	0	0	0	1	0	0
XII	10	5	2	0	0	1	1	2	1	2	2	2	0	0
XIII	2	0	0	2	0	0	1	2	0	1	0	1	0	0
E	0	0	0	0	0	0	0	0	0	1	0	0	0	0

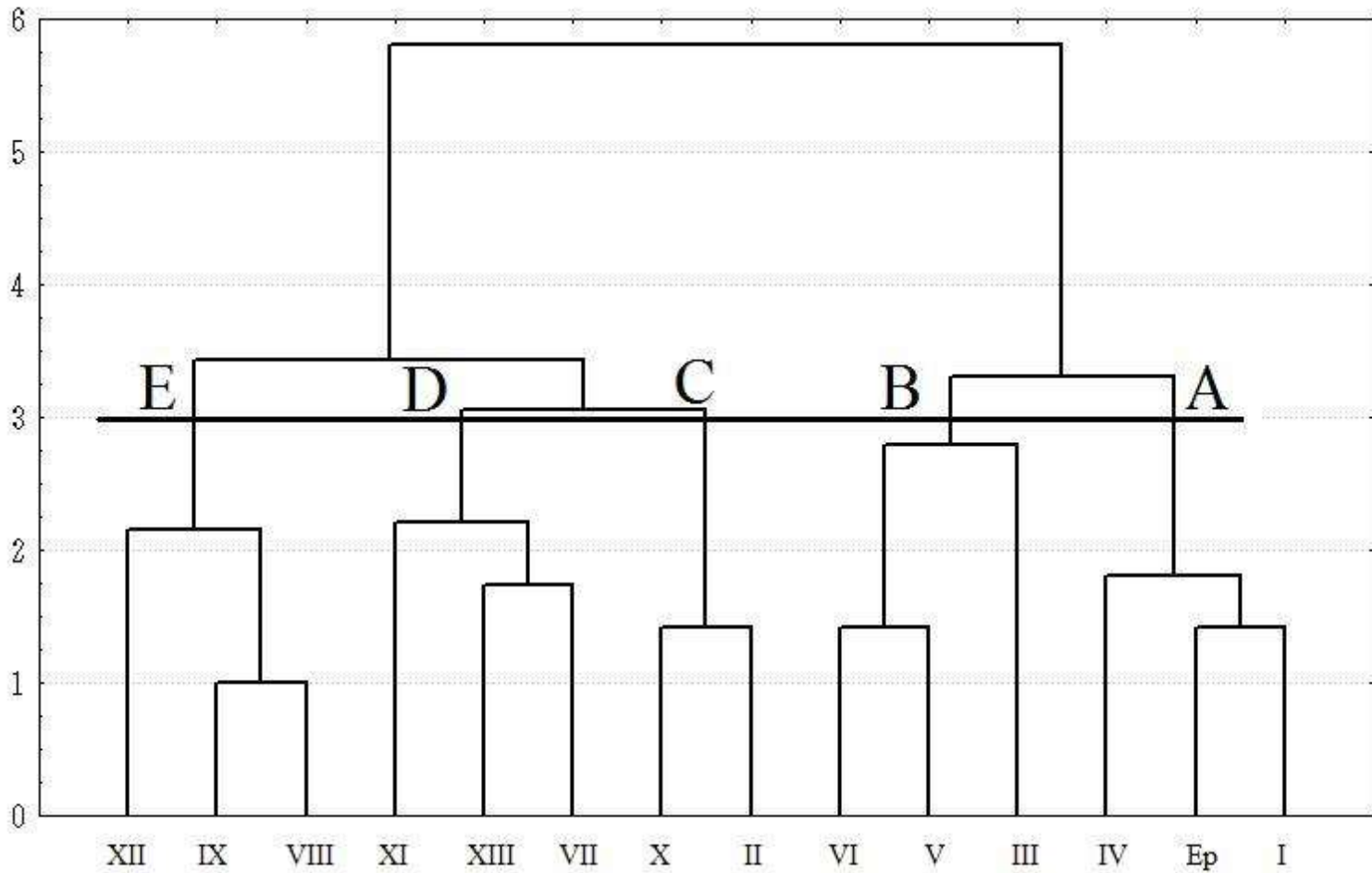
Matrice de données pondérées



	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
I	0	0	0	0	0	0	0	0	0	0	0	1	0	0
II	1	1	0	0	1	0	1	1	0	1	0	0	0	1
III	0	1	0	0	1	0	0	0	1	1	1	0	1	1
IV	1	0	0	0	0	0	0	0	0	1	0	0	0	1
V	0	1	0	0	0	0	1	0	0	0	0	0	0	1
VI	0	1	1	0	0	0	1	0	0	0	0	0	1	1
VII	1	0	0	0	0	0	1	0	1	1	0	1	0	0
VIII	1	1	1	1	0	0	1	1	0	1	1	1	1	0
IX	1	1	1	1	0	0	1	1	0	1	1	1	0	0
X	1	1	0	0	0	0	1	1	0	1	0	1	0	1
XI	1	0	1	1	0	1	1	1	0	0	0	1	0	0
XII	1	1	1	0	0	1	1	1	1	1	1	1	0	0
XIII	1	0	0	1	0	0	1	1	0	1	0	1	0	0
E	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Matrice de données dichotomiques

Analyse cluster

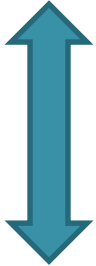


SPSS version académique 18.0.0 (2009)
Dendrogramme de l'analyse cluster de quatorze textes,
Méthode Ward, Distance euclidienne



dendron (δένδρον) = arbre

Hiérarchie des indicateurs graphiques

Regroupement des textes de TM	Indicateurs graphiques	
Groupe A (I, IV, Ep)	aucun indicateur	sans régionalisme  régionalisme bourguignon
Groupe B (III, V, VI)	(2) –aige, (14) consoil, mervoille, etc	
Groupe C (II, X)	(1) confusion entre ai, e et a (8) chute de –s-, (10) arme	
Groupe D (VII, XI, XIII)	(4) –auble, (12) virge, virges	
Groupe E (VIII, IX, XII)	(3) –arent, (11) yglise	

Corrélation entre 5 groupes

	GroupeA	GroupeB	GroupeC	GroupeD	GroupeE
GroupeA	0.000	2.906	3.298	7.984	*12.911
GroupeB	2.906	0.000	0.073	1.928	4.325
GroupeC	3.298	0.073	0.000	1.040	2.361
GroupeD	7.984	1.928	1.040	0.000	0.120
GroupeE	*12.911	4.325	2.361	0.120	0.000

Matrice de distance de χ^2

* = significatif; $p \leq 0.05$

La distance entre Groupe ABC (I, Épilogue, IV, III, V et VI) et Groupe DE (II, X, VII, XIII, XI, VIII, IX et XII) est hautement supérieure à la distance critique: $\chi^2 = 9.953$, degré de liberté = 1, $p = 0.001606$.

Réponse aux questions

- Il est difficile d'admettre que *TM* est un texte linguistiquement homogène vu la micro-variation de graphies.
- Il serait aussi hasardeux de prétendre que *TM* est un texte bourguignon, puisque les premier et quatrième miracles et l'épilogue en particulier, sont des textes sans traits régionaux nets.



3. Graphies médiévales et changements phonétiques en cours (1230-1300)

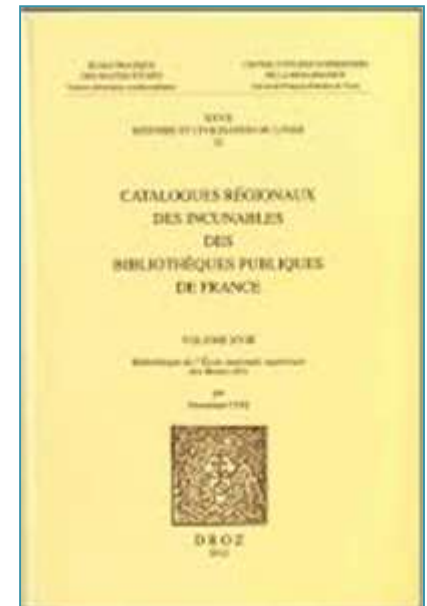
Questions de recherche

Les graphies dans les actes ont-elles pour fonction d'indiquer des changements phonétiques qui se déroulent dans le temps et l'espace ?

Corpus

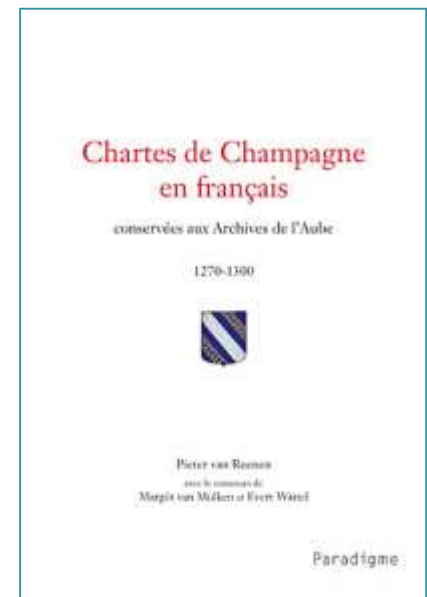
Dominique Coq (1988)

Documents Linguistiques de la France (série française), III, *Chartes en langue française antérieures à 1271 conservées dans les départements de l'Aube, de la Seine-et-Marne et de l'Yonne*, Paris, Éditions du CNRS. plus de 45000 mots.



Pieter van Reenen (2007)

Chartes de Champagne en français conservées aux Archives de l'Aube, 1270-1300, avec le concours de Margôt van Mulken et Evert Wattel, Orléans, Paradigme. 42000 mots.



Fonction indicatrice chronologique des graphies

latin : *diurnu*



ancien français : *jor*



français moderne : *jour*

Passage de -o- [o] vers -ou- [u]

ancien français

cort, jor, tot, etc.



XII^e siècle -ou-

court, jour, tout, etc.

Eviter l'erreur type

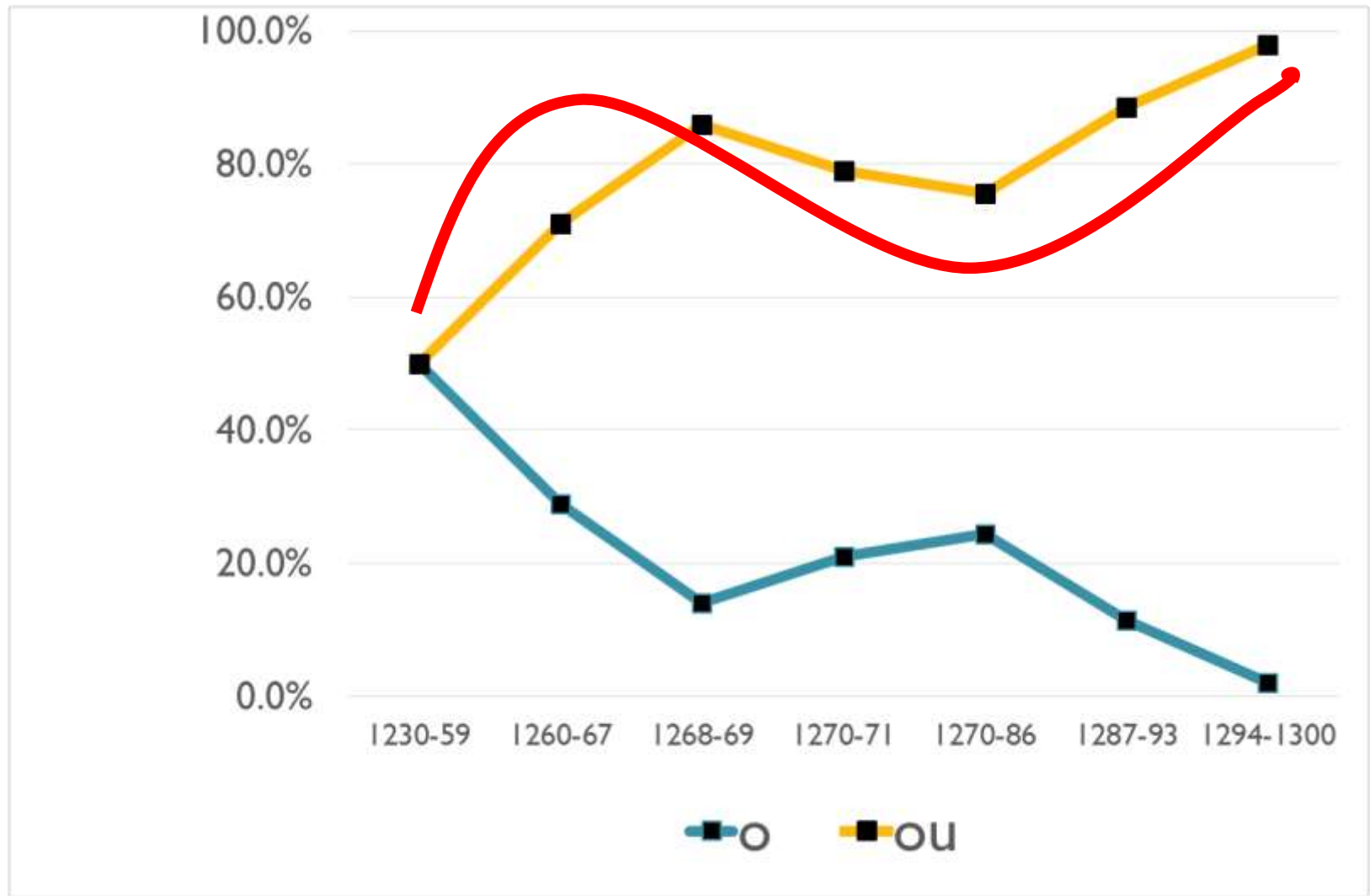
Concordance Hits 5

Hit	KWIC	File
1	gneor Guill/aum/e de La- Cort , lou chevalier, /et/ es	004.txt
2	sunt seeles ou seel de la cort de Troies de ce don ,	036.txt
3	vant en plait, en quelque cort que ce fust, /et/ qui a	054.txt
4	ou p/ar/ jugement en-sa- cort 101[F3/17R] puis qu	101.txt
5	~ dele~s la 102[F63/2R] Cort Nostre Dame /et/ p/	102.txt

Concordance Hits 8

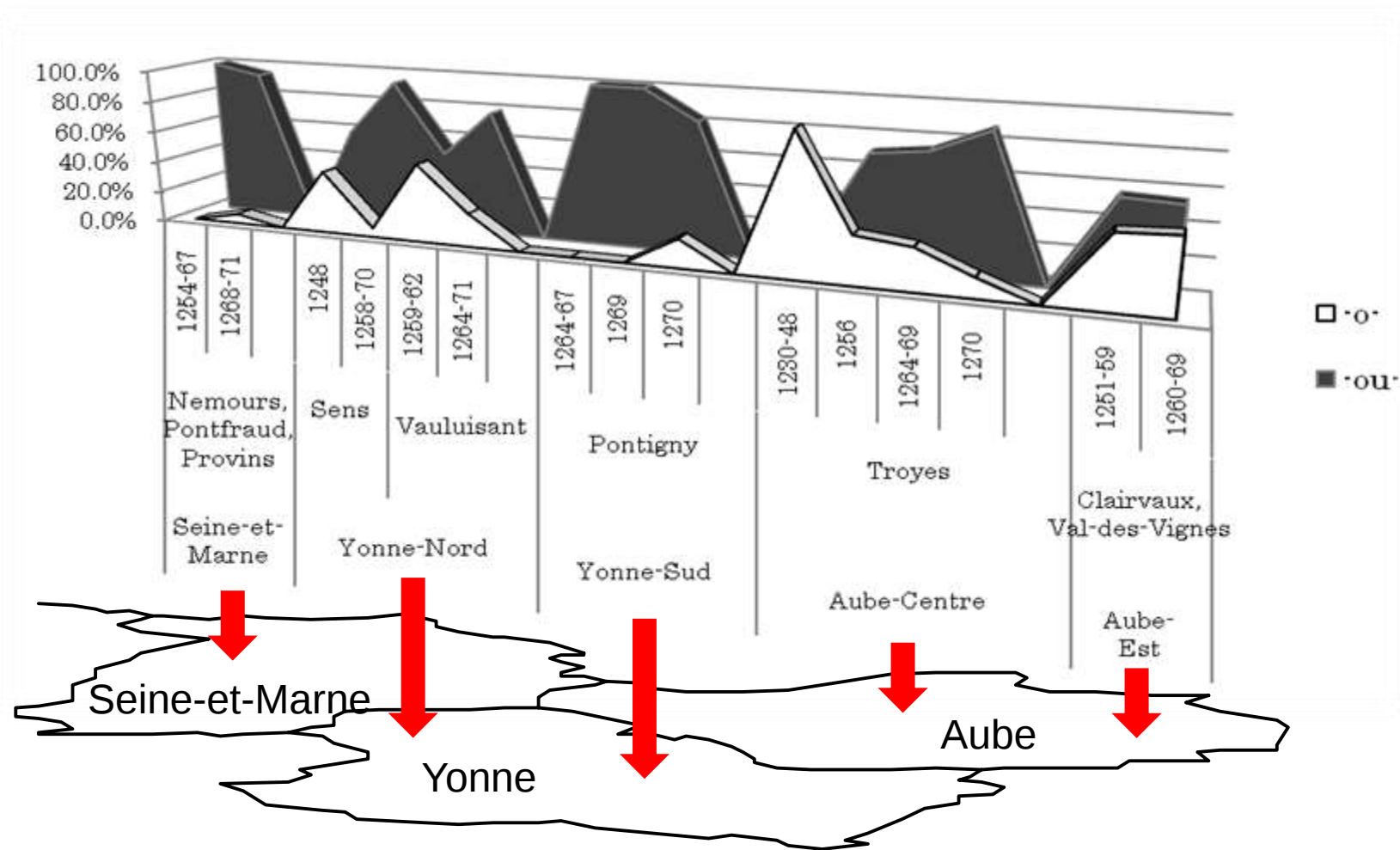
Hit	KWIC	File
1	andement l-official de la court de Sanz tout le droi	041.txt
2	eesse e au couvent de la Court Nostre Dame delez	050.txt
3	esse /e/ au-covant de-la- Court Nostre Dame /XX/	091.txt
4	ez je lou a-l-abesse de-la- Court Nostre 091[4] Dam	091.txt
5	er/res as nonay/n/s de-la- Court Nostre Dame /e/ la	091.txt
6	esse /e/ au-covant de-la- Court Nostre Dame co/m	091.txt
7	esse /e/ au-covant de-la- Court Nostre Dame bien ,	091.txt
8	sse /et/ au couvant de la Court Nostre Dame delez	099.txt

Évolution de -o- > -ou-



*jor, jour, jors, jours, jorz, jourz, touzjourz,
tot, tout, tote, toute, toz, touz, etc.*

Variation géolinguistique entre -o- / -ou-



Évolution de -o- > -ou- selon les régions

Réponse à la question de recherche

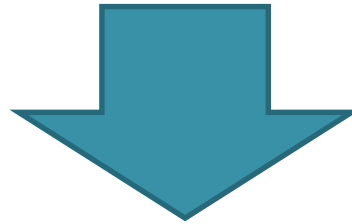
- Nous avons pu retracer les étapes successives de l'évolution de la graphie
-o- en -ou-, qui avait commencé au début du XIII^e siècle dans la Champagne occidentale et qui s'est propagée ensuite dans toute la Champagne vers la fin du siècle. En bref, les graphies médiévales ont pour fonction d'indiquer le changement dans le temps et l'espace.



4. Témoignages du français parlé au XVIIIe siècle

Question de recherche

Peut-on étudier la langue parlée dans les textes écrits ?

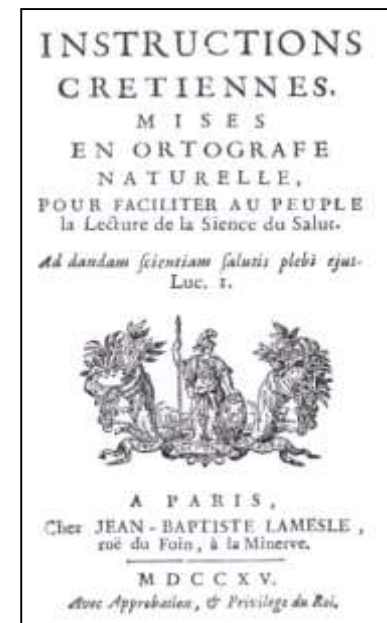
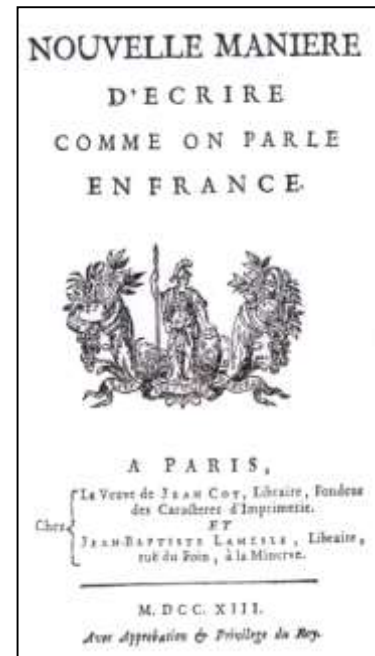


Oui, car il y a des textes qui semblent refléter la langue parlée de l'époque.

Gile Vaudelin

Nouvelle manière d'écrire comme on parle en France, 1713, Paris: Chez La Veuve de Jean Cot et Jean-Baptiste Lamesle, Slatkine Reprints, 33p. (NM)

Instructions crétiennes, mises en ortografe naturelle, pour faciliter au peuple la lecture de la Sience du salut, 1715, Paris: Chez Jean-Baptiste Lamesle, Slatkine Reprints, 247p. (IC)



Alphabet nouveau

ALPHABET NOUVEAU.

a. an. ai. é. in. i. e. o. on. eu. un. ou u. valeur.
A. A. A. E. I. I. E. O. O. E. U. O. U. } figures.
a. a. a. e. i. i. e. o. o. e. u. o. u. }

que. gue. che. je. ne. le. re. ze. se. de. te. valeur.
C. G. H. J. N. L. R. Z. S. D. T. } figures.
c. g. h. j. n. l. r. z. f. d. t. }

ve. fe. pe. be. me. valeur.
V. F. P. B. M. }
v. f. p. b. m. } figures.

13 voyelles

16 consonnes

Corpus et transcription

Reflacsiq.

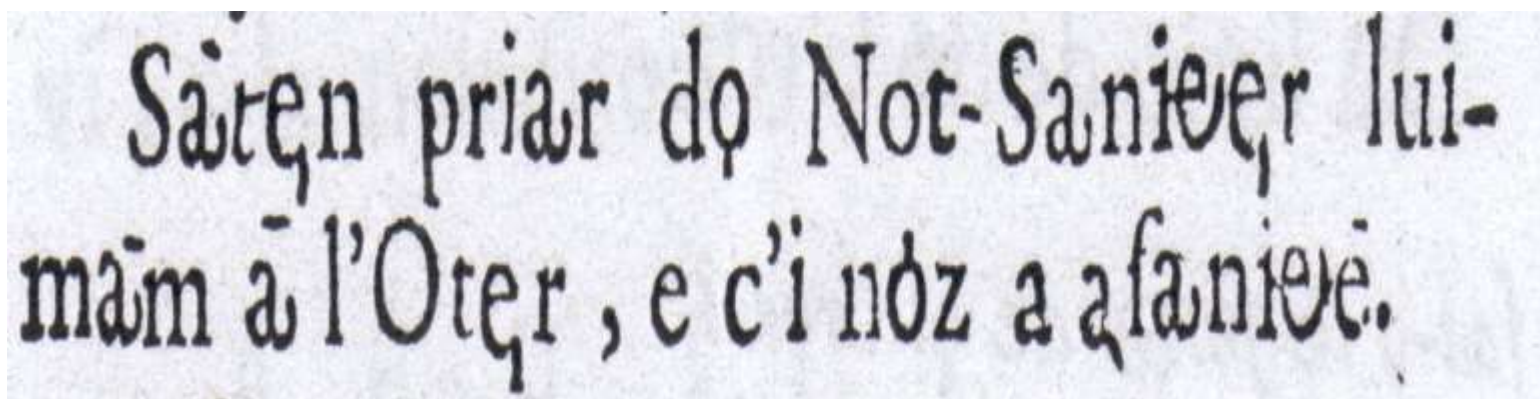
1. Si d'Ecrire ôtrma ce l'ô Parl a
Fras, i n'an ariva ce pe, ô ce de pûz
icoveniã; e si l'ô ne s'a plainiea po-i
partô, e dpui lq-ta, parson n'ora
jama pase a la Reform de l'Ortograf
Frasâz.

Rflaicsion.

1. Si d'Ecrire ôtrman ce l'on Parl an
Frans, i n'ann arivai ce peu ou de ptîz
inconveniân; e si l'on ne s'an plainieai po-in
partou, e dpui lon-tan, pairson n'orai
jamai panse a la Reform de l'Ortograf
Fransâiz.

NM, p. 17

Liaison du verbe être



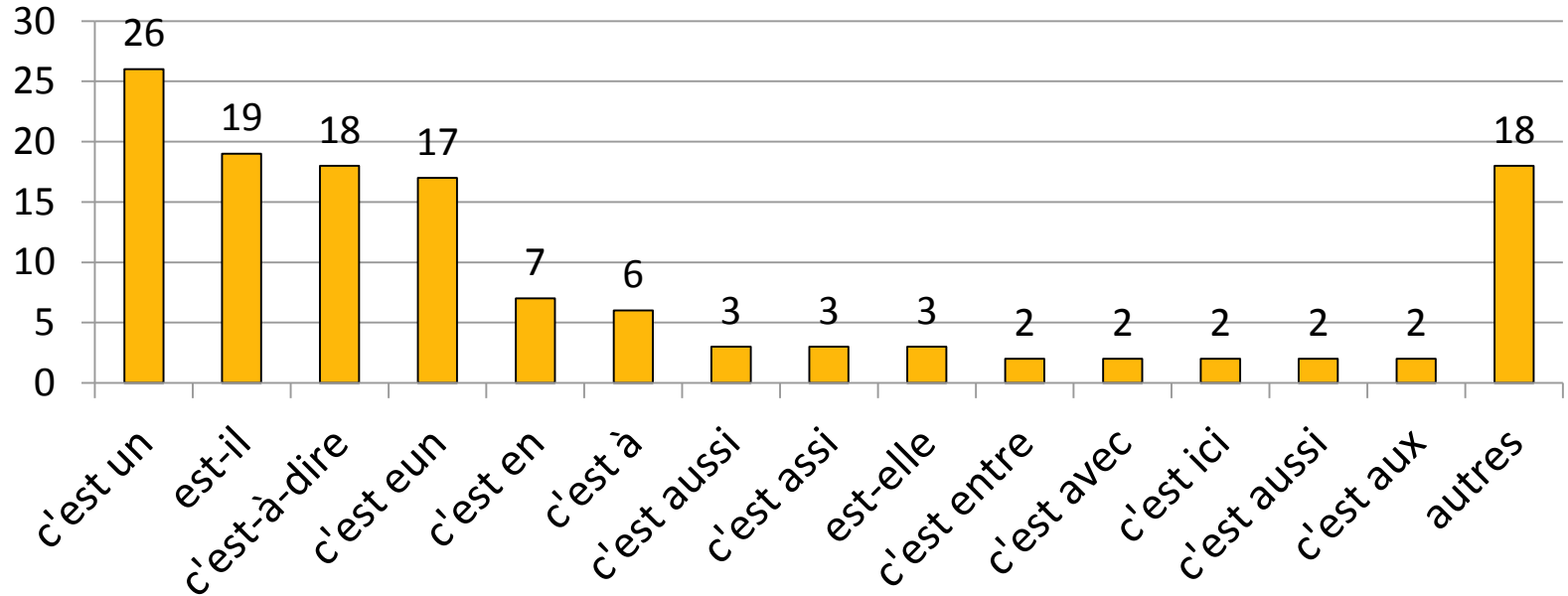
Sâiteun priar do Not-Sainieer lui-
mâm â l'Oteur, e c'i nòz a a fanieê.

Sâiteun priair don Not-Sainieeur lui-
mâim âi l'Oteur, e c'i nouz a ansainieê.

= C'est une prière dont Notre-Seigneur lui-
même est l'auteur, et qui nous a enseignée.

IC, p. 138

Liaison du verbe être



Distribution de “*est*” en liaison en occurrences

C'est sans liaison:

Pourcoai dit-vou ce *s'âi eun* espri? IC, p.81.
= Pourquoi dites-vous que c'est un esprit?

Réponse à la question de recherche

- À condition que les textes semblent refléter la langue parlée de l'époque, on peut étudier la langue parlée dans les textes écrits.



5. Sociolinguistique et pragmatique du japonais au début XIXe siècle.

Question de recherche

Peut-on étudier la langue parlée du siècle passé comme la langue actuelle, du point de vue sociolinguistique par exemple ?

3 .*Oukiyoubouro* 『浮世風呂』

Shikitéi Samba 式亭三馬

Tome I (1809) 20 scènes

Tome II (1810) 12 scènes

Tome III (1812) 16 scènes

Tome IV (1813) 15 scènes



Rakougo
“Histoire comique
racontée”



Variation de la forme négative

Auxiliaire négative **-nou** / **-nai** / **-née**

mi-**nou** / mi-**nai** / mi-**née**

“(je) ne regarde pas”

Terminaison adjectivale de **-ai** / **-ée**

aboun-**ai** / aboun-**ée** “dangereux”

tchiis-**ai** / tchiis-**ée** “petit”

Différence sexuelle

45 hommes et 53 femmes

	moyenne d'occurrences				
	masculine	féminine	t	df	p
-nou	1.844	0.434	2.723	96	0.008 **
-naï	0.067	0.113	-0.707	96	0.481
-née	3.756	3.585	0.191	96	0.849
-aï	10.644	10.774	-0.047	96	0.963
-ée	9.378	6.849	1.018	96	0.311

terminaison masculine **-nou** ?

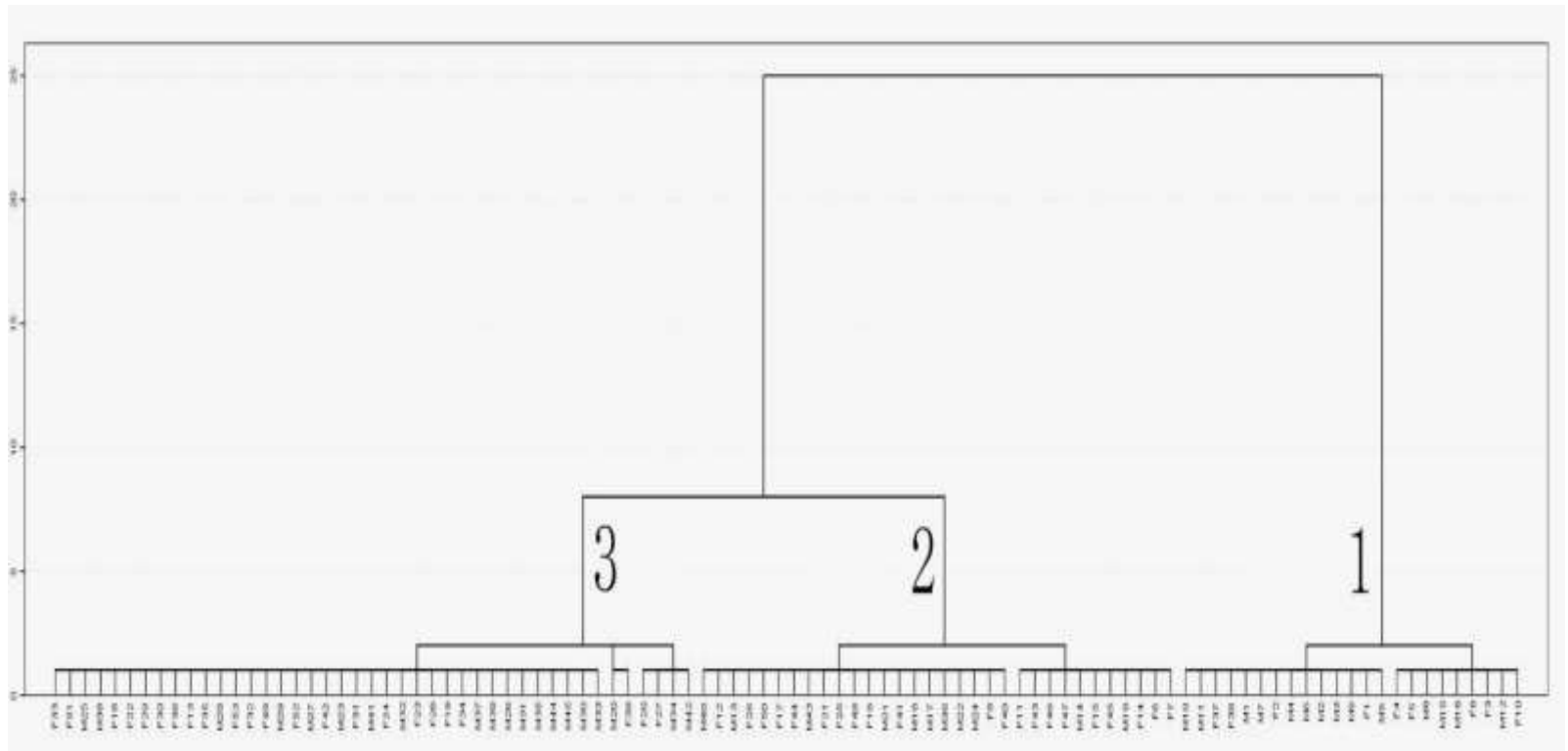
cf. masculin yatta-zo “(j)’ai fait”

féminin yatta-wa

Matrice du pourcentage d'utilisation de cinq variantes (extrait)

	Personnages	Sexe	-nu	-nai	-nê		-ai	-ê
M1	Kaki no itchi	M	100%	0%	0%		100%	0%
M2	Inkyo	M	100%	0%	0%		100%	0%
:	:	:	:	:	:		:	:
M45	Katsubê	M	0%	0%	100%		6.4%	93.6%
F1	Oyasu	F	100%	0%	0%		98%	2%
F2	Kamoko	F	100%	0%	0%		100%	0%
:	:	:	:	:	:		:	:
F53	Osaru	F	0 (0%)	0 (0%)	100%		21.4%	78.6%

Analyse de cluster



SPSS version académique 18.0.0 (2009)

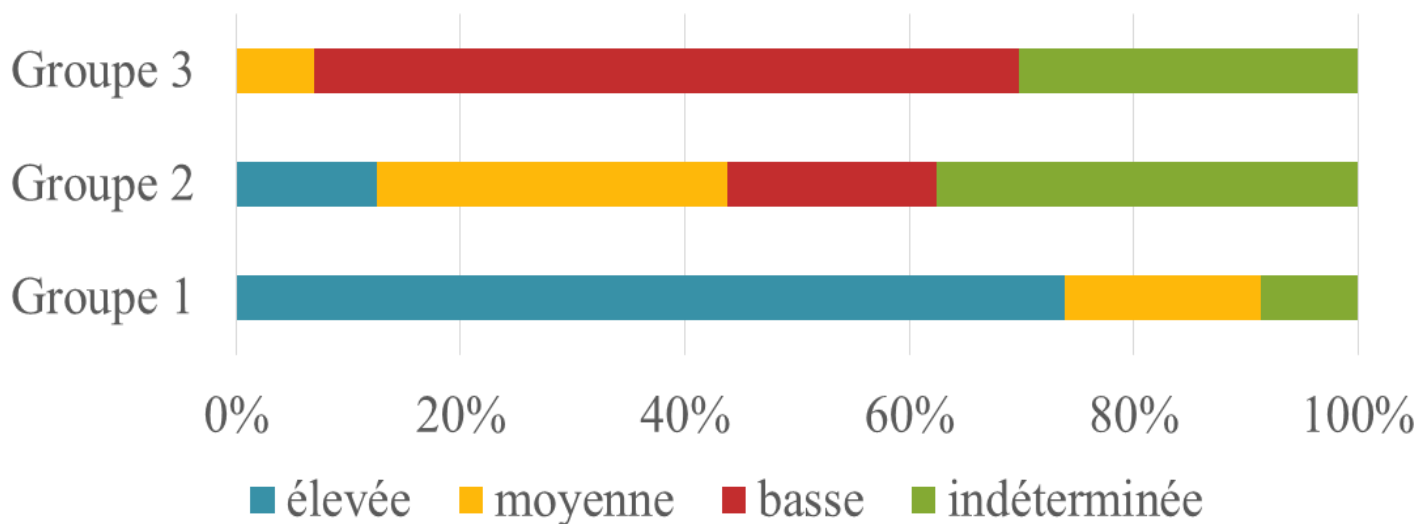
Dendrogramme de groupements des 98 personnages

Méthode Ward, Distance euclidienne

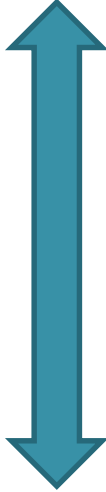
Groupes majeurs et classes sociales

Classe sociale	Groupe 1 23 personnes	Groupe 2 32 personnes	Groupe 3 43 personnes
Élevée	17 pers.	4 pers	
Moyenne	4 pers.	10 pers.	3 pers.
Basse		6 pers.	27 pers.
Indéterminée	2 pers.	12 pers.	13 pers.

Constitution des trois groupes majeurs

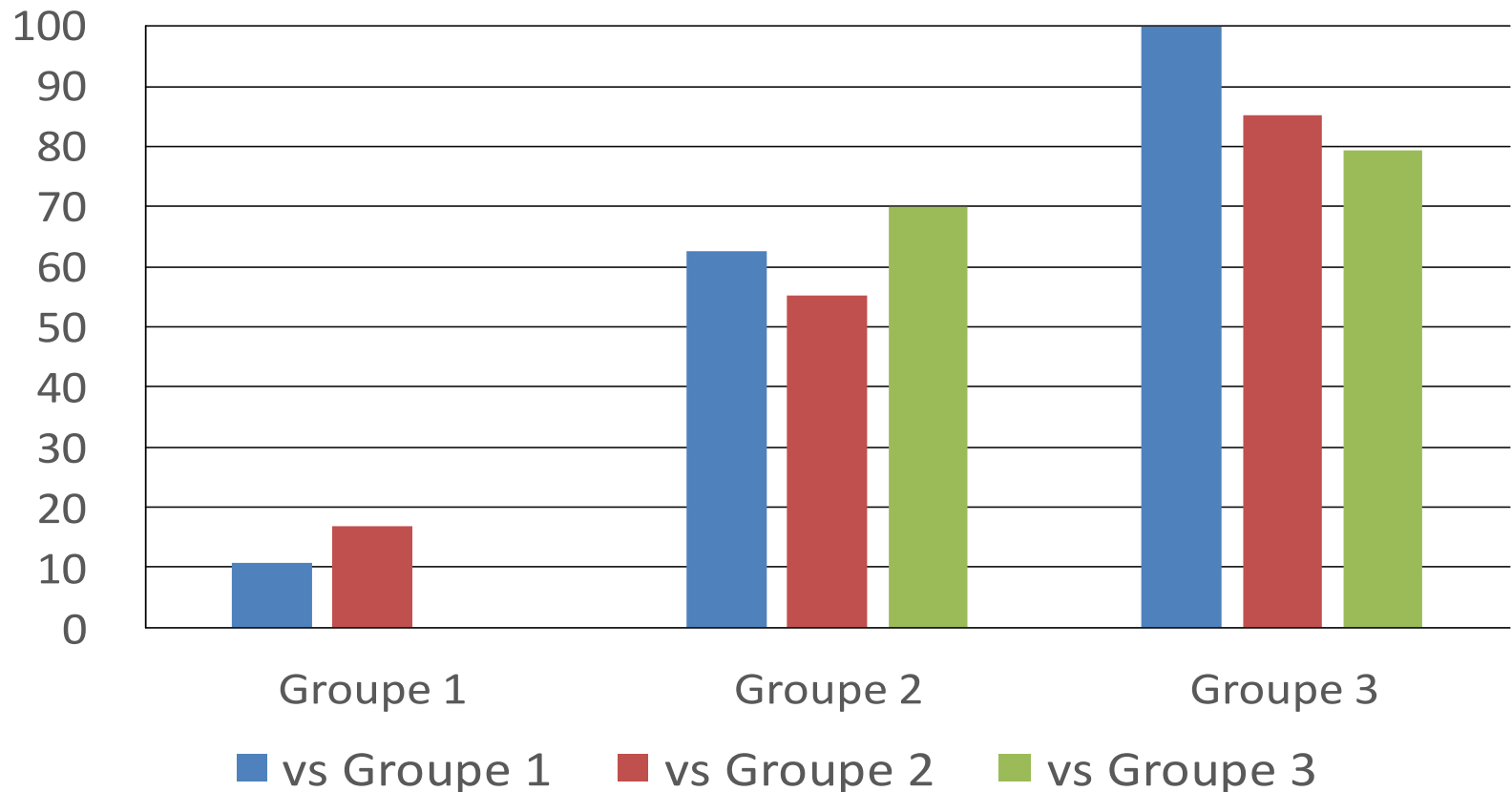


Groupes majeurs et variantes

	Hommes	Femmes	Statut social	Variantes
Groupe 1	Inkyo «retraité» Isha «docteur» Kikaku dit «maître» Namayoi samourai», etc.	Oyasu, Otatsu «épouses de commerçants», etc.	 Élevé	-nu et -ai prédominantes
Groupe 2	Mudanosuke, Tobihachi «hommes débauchés» Bantô «commis», etc.	Okako, Osami «geishas», etc.		-nê et -ai prédominantes
Groupe 3	Yukumi «serveur de bain» Isami, Teppôsaku «hommes débauchés», etc.	Omaru «bonne» Oshita, Otobi «filles publiques» Otai «serveuse», etc.		-nê et -ê prédominantes

Variation pragmatique chez les locuteurs

Pourcentage d'utilisation de «-ê» dans les trois groupes



Réponse à la question

- On peut étudier la langue parlée du siècle passé comme la langue actuelle
- On peut même faire une analyse variationnelle de langue parlée dans les textes écrits

Pour conclure

Au bout de ce long voyage intellectuel:

- Pour réaliser l'analyse linguistique de la variation des formes dans les textes de siècles passés:

Disponibilité de divers paramètres analytiques:
philologie et scriptologie médiévales, dialectologie,
phonétique historique et sociolinguistique, etc.

- Pour se rendre compte de phénomènes dans leur ampleur et profondeur:

Approche combinée de la linguistique du corpus et de la
méthode statistique

Références

- Cohen, Marcel. 1946. *Le français en 1700 d'après le témoignage de Gile Vaudelin*, Paris: Librairie ancienne Honoré Champion.
- Coq, D. (1988). Documents Linguistiques de la France (série française), III, Chartes en langue française antérieures à 1271 conservées dans les départements de l'Aube, de la Seine-et-Marne et de l'Yonne, Paris, Éditions du CNRS.
- Dees, A. (1980). (avec le concours de P. Th. van Reenen et de Johan A. de Vries) Atlas des formes linguistiques des chartes françaises du 13e siècle, Tübingen, Max Niemeyer Verlag.
- Durand, Jacques et Chantal Lyche. 2008. "French liaison in the light of corpus data", *French Language Studies*, 18, 33-66.
- Fouché, Pierre (1969). *Phonétique historique du français, Volume II Les Voyelles*, 2^e édition, Paris, Klincksieck.
- Kawaguchi, Y. (2005). *Recherches linguistiques sur le champenois méridional au moyen âge -Aspects phonétiques et graphiques-*, Lille, Atelier national de reproduction des thèses.
- (2007). L'État actuel de la dialectologie du français médiéval: le cas des chartes champenoises méridionales, *Le Nouveau Corpus d'Amsterdam*, Zeitschrift für französische Sprache und Literatur - Beihefte Band 34, P. Kunstmann et A. Stein éd., Stuttgart: Steiner, 188-200.
 - (2008). Micro-variation de graphies dans Treize Miracles de Notre-Dame, *Por s'onor croistre Mélanges de langue et de littérature médiévales offerts à Pierre Kunstmann*, Les Éditions David, 169-180.
 - (2008). Sur les fonctions indicatrices chronologique et géolinguistique des scriptas médiévales : le cas des actes champenois méridionaux (1230-1300), 2008, *Le Premier Congrès Mondial de la Linguistique française*, le 9 juin, 2008, Paris.
 - (2010). French Liaison in the 18th Century -Analysis of Gile Vaudelin's texts-, *Corpus-based Analysis in Diachronic Linguistics*, Y. Kawaguchi, M. Minegishi, W. Viereck (eds.), John Benjamins, 133-151.
 - (2012). Peut-on étudier la langue parlée dans les documents écrits ? Témoignages du français et du japonais, *Penser les langues avec Claire Blanche-Benveniste*, Sandrine Caddéo, Marie-Noëlle Roubaud, Magali Rouquier et Frédéric Sabio (dir). Presses Universitaires de Provence, 235-247.
- (2015). Formule d'adresse et variation entre *faire savoir* et *faire a savoir* dans les actes champenois au XIII^e siècle, *Etudes de linguistique médiévale. Hommage à Ambroise Jean-Marc Queffélec 1*, Peter Lang, 93-108.
- Nakazawa Noriko (2006) "Opposition entre *nu* et *nee* dans l'auxiliaire négative dans la langue de Edo", (en jp.) *Nihongo no kenkyû* 2.2., pp.93-107.
- Nakata Toshio (1985) "Changement de langage suivant les situations langagières dans *Oukiyoubouro* – Avec le témoignage de l'alongement vocalique de *-ai-*", (en jp) *Jinbun gakuho* 173, pp.25-47.
- Monfrin, J. (1968). «Le mode de tradition des actes écrits et les études de dialectologie», *Revue de Linguistique Romane*, 32, 18-47.
- Pope M.K. (1973). *From Latin to Modern French with especial consideration of Anglo-Norman phonology and morphology*, Reprinted, Manchester, Manchester University Press.
- Reenen, P. Th. v. (2007). Chartes de Champagne en français conservées aux Archives de l'Aube, 1270-1300, avec le concours de Margôt van Mulken et Evert Wattel, Orléans, Paradigme.

AntConc : <http://www.laurenceanthony.net/software.html>

Remerciements

These different researches are supported by

JSPS KAKENHI Grant Numbers 24320102

Grant-in-Aid for Scientific Research(B)

Responsible: Yuji KAWAGUCHI.