

Langue parlée dans les documents écrits

– Cas du français et du japonais –

Yuji KAWAGUCHI

Tokyo University of Foreign Studies (TUFS)

**«Penser les langues avec Claire Blanche-Benveniste»
2-4 décembre, 2010, Paris**

Plan

1. Caractéristiques du français parlé
2. Français parlé dans les textes écrits
3. Japonais parlé dans les textes écrits

Conclusion

Pour commencer

What's New
(218966 words)

Outline & Purpose
CbLLE POS Research Engine (Spoken French) is to search part-of-speech tags of the *Multilingual Spoken Language Corpora (French)* developed by the 21st Century COE Program "Usage-Based Linguistic Informatics". Part-of-speech (POS) tagging is based on the system called *Tree Tagger*, cf. <http://www.lms.uni-stuttgart.de/projekte/complex/TreeTagger/>. The purpose of this research engine is largely to publish the research results from the Global COE Program "Corpus-based Linguistics and Language Education (CbLLE) and to analyze various grammatical phenomena in spoken French.

Corpus Description
The current corpora of Spoken French are composed of two different datas.
1. Twenty dialogs recorded during July 2005 with the cooperation of the University of Aix-Marseille I. The total length is around seven hours and the total number of word tokens exceeds 100,000. They cover various subjects such as office life, work at the university, recent news about one's family and improvised sketches.

References
http://www.coelang.tufs.ac.jp/multilingual_corpus/fr/index.php?contents_xmFgaisetsu&lang=ja
http://www.coelang.tufs.ac.jp/multilingual_corpus/fr2/index.php?contents_xmFgaisetsu&lang=ja

Research Team
Supervisors:
- Atsushi Sano (Professor of Fukushima University)
- Yuji Kawasuchi (Professor of Tokyo University of Foreign Studies)
Organizers:
- Christophe Beniztoun (Post-doctoral fellow of University of Aix-Marseille)
- Hisae Akhino (Ph.D. of École Pratique des Hautes Études)
- Françoise Lorant (University of Paris XIII)
- Sunshuke Nakata (Ph.D. Candidate for Tokyo University of Foreign Studies)
- Mito Matsuzawa (MA Tokyo University of Foreign Studies)
- Takahiro Ogasawa (MA Tokyo University of Foreign Studies)
Academic Advisers:
- André Valli (Professor of University of Aix-Marseille)
- José Delofeu (Professor of University of Aix-Marseille)
- Claire Blanche-Benveniste (Professor emeritus of University of Aix-Marseille)

Recherches: Corpus du français parlé
Depuis 2005, 218,000 mots;
Ettiueté avec Tree-Tagger

Pôle d'excellence du Japon:
Corpus-based Linguistics and Language
Education, Tokyo University of Foreign
Studies avec la collaboration de Université
de Aix-Marseille

<http://cbllle.tufs.ac.jp/tag/fr/index.php?menulang=en>

pos: Any:POS Any:POS Any:POS Search 0 0 Français ja HELP search sample

lemma: on Reset par: [x] aix: [x]

token: [] Ignore the second context

Conserve results Conserve results with contexts

No of sentences :1630
No of tokens :2186

- fr57.txt:10 on est où là déjà c' est euh
- fr57.txt:153 j' ai perdu mon sac on m' a volé mon sac +
- fr57.txt:182 mais s' il vous plaît > aidez -moi on m' a volé mon sac +
- fr57.txt:196 on m' a volé mon sac +
- fr57.txt:230 non j' ai plus rien on m' a volé toutes mes affaires je sais pas du tout où je suis +
- fr57.txt:369 alors quand on est adolescent ça va ça passe quoi mais bon après euh non +
- fr57.txt:410 ce que j' étais réellement euh à l' époque où on s' est rencontré hein +
- fr57.txt:414 juste à ce qu' on va saaner euh +
- fr57.txt:415 aux impôts qu' on va payer euh c' est pas ça l' important dans la vie +
- fr57.txt:416 l' important c' est de s' amuser d' être heureux et de faire ce qu' on aime non tu crois pas +
- fr57.txt:472 ben enfin je veux dire on aura beau t' enfermer pendant vingt ans dans une cave ou dans cette / salle là , salle -là / euh +

1. Caractéristiques du français parlé

15 traits, Blanche-Benveniste et Jeanjean (1987)

14 traits, Koch et Oesterreicher (1990)

8 traits morpho-syntaxiques communs

(1) usage de *ça* ;

(2) [i] pour *il* ;

(3) non-usage de négative *ne* ;

(4) futur périphrastique *aller + inf.* ;

(5) disparition du subjonctif;

(6) déclin de l'inversion interrogative;

(7) reprise du pronom clitique;

(8) *on* pour *nous*;

2. Français parlé dans les textes écrits

Le 21 septembre 1601

“sur quatre heures du soir, à l’entrée du jardin des canaux, je rencontrai le Roi qui revenoit de la chasse, et m’appelant, me fit l’honneur de me dire: «Je vous ai choisi pour vous mettre près de mon fils le Dauphin; servez-le bien».”



Louis XIII

2.1. Journal de Jean Héroard (1605-1610)

cf. Ernst (1985)

1. Non-usage de *ne*

avec verbes *oser, savoir, pouvoir*

je pui pas resoudre cela 18.10.1610

2. Inversion interrogative: Exemples attestés en 1608

Inversion (53 occ.)

est ce assés? ay je pa la mienne? quel aage atelle?
son frere eti saint? ou allé vou?

Non-inversion (3 occ.)

mais jl est en paradi? que je li ay rien envoié?
papa e pa encore venu?

2.2. Agréable Conférence de deux Normands (1652)

cf. Bougy (1996)

je veux que la **basse dichin** nous **feche** demein de la marende, parche que je ne serez macquis de **pous**, on bien du tout il faut dire à la **basse** que nous **feiche** des **Craules**, (l.75-77)

“je veux que la **servante d’ici** nous **fasse** demain une collation, parce que je ne saurais manger de la **bouillie**. Ou bien, oui, il faut dire à la **servante** qu’elle nous **fasse** des **craules**,” cf. Bougy (1996)

Variation diatopique

Changements linguistiques en cours

Inversion interrogative

	Inversion	Non-inversion
Journal de J. Héroards, l'an 1608	53 occ. (95%)	3 occ. (5%)
Agréable Conférence (A.C.), 1652	8 occ. (47%)	9 occ. (53%)

vous ne sçavais pay qui lond tout ravagié les gardins?

(A.C.I.5)

“Vous ne savez pas qu'ils ont tout ravagé les jardins?”

Les parisiaux estee donque bien effretais de ves chela?

(A.C I.9)

“Les Parisiens étaient donc bien effrayés de voir cela?”

3 . Japonais parlé dans les textes écrits

Oukiyobouro 『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 du japonais parlé

Auxiliaire négative **-nou** / **-naï** / **-née**

mi-**nou** / mi-**naï** / mi-**née**

“(je) ne regarde pas”

Terminaison de **-aï** / **-ée**

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

tchiis-**aï** / tchiis-**ée** “petit”

45 hommes et 53 femmes

cf. Nakazawa (2006)

Différence sexuelle

	moyenne d'occurrences				
	masculine	féminine	t	df	p
-nou	1.8	0.434	2.72	96	0.008 < 0.05 (S)
-naï	0.1	0.113	-0.7	96	0.481
-née	3.8	3.585	0.19	96	0.849
-aï	11	10.77	-0	96	0.963
-ée	9.4	6.849	1.02	96	0.311

terminaison masculine **-nou** ?

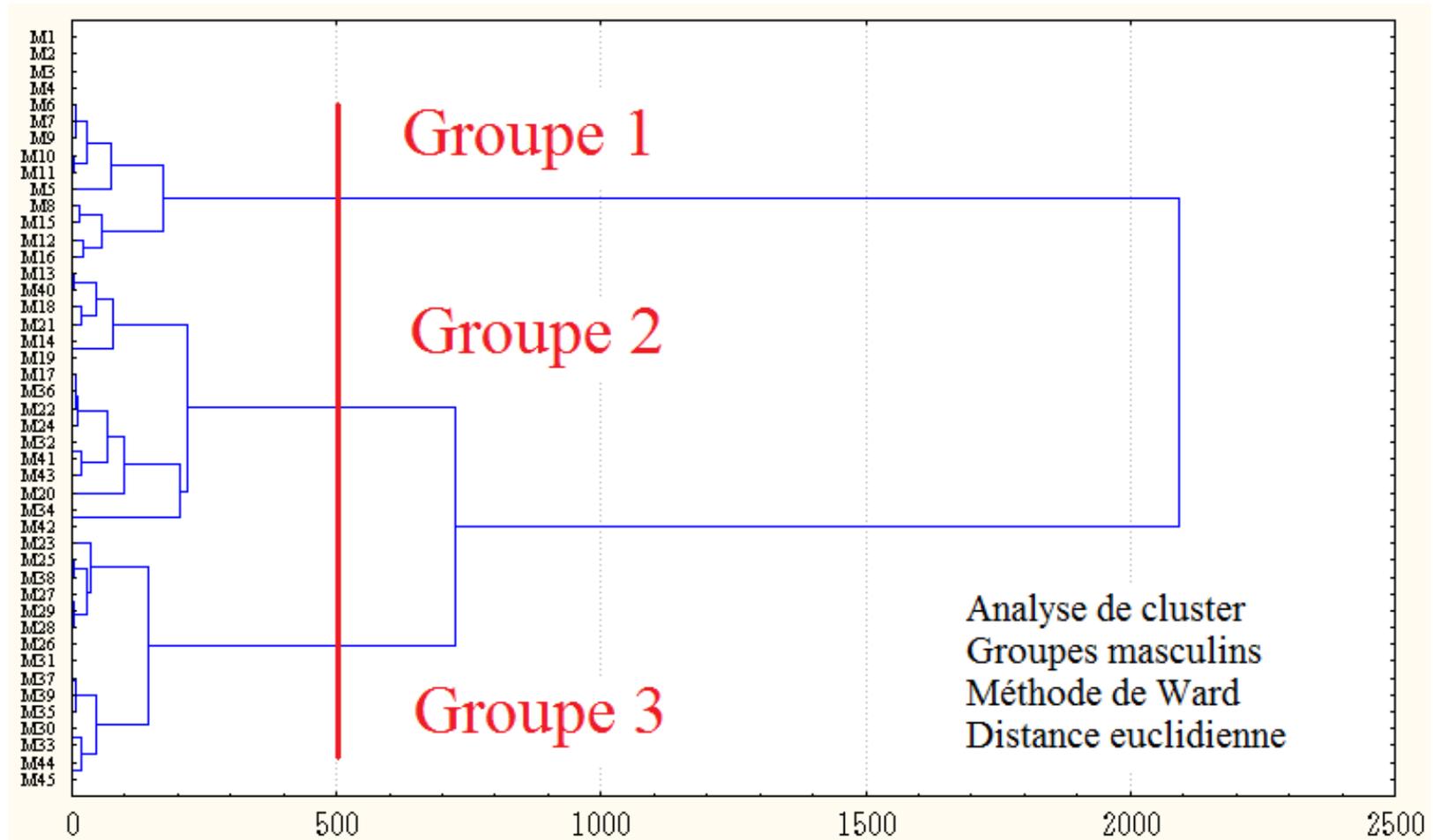
cf. masculin yatta-*zo* “(j)’ai fait”

féminin yatta-*wa*

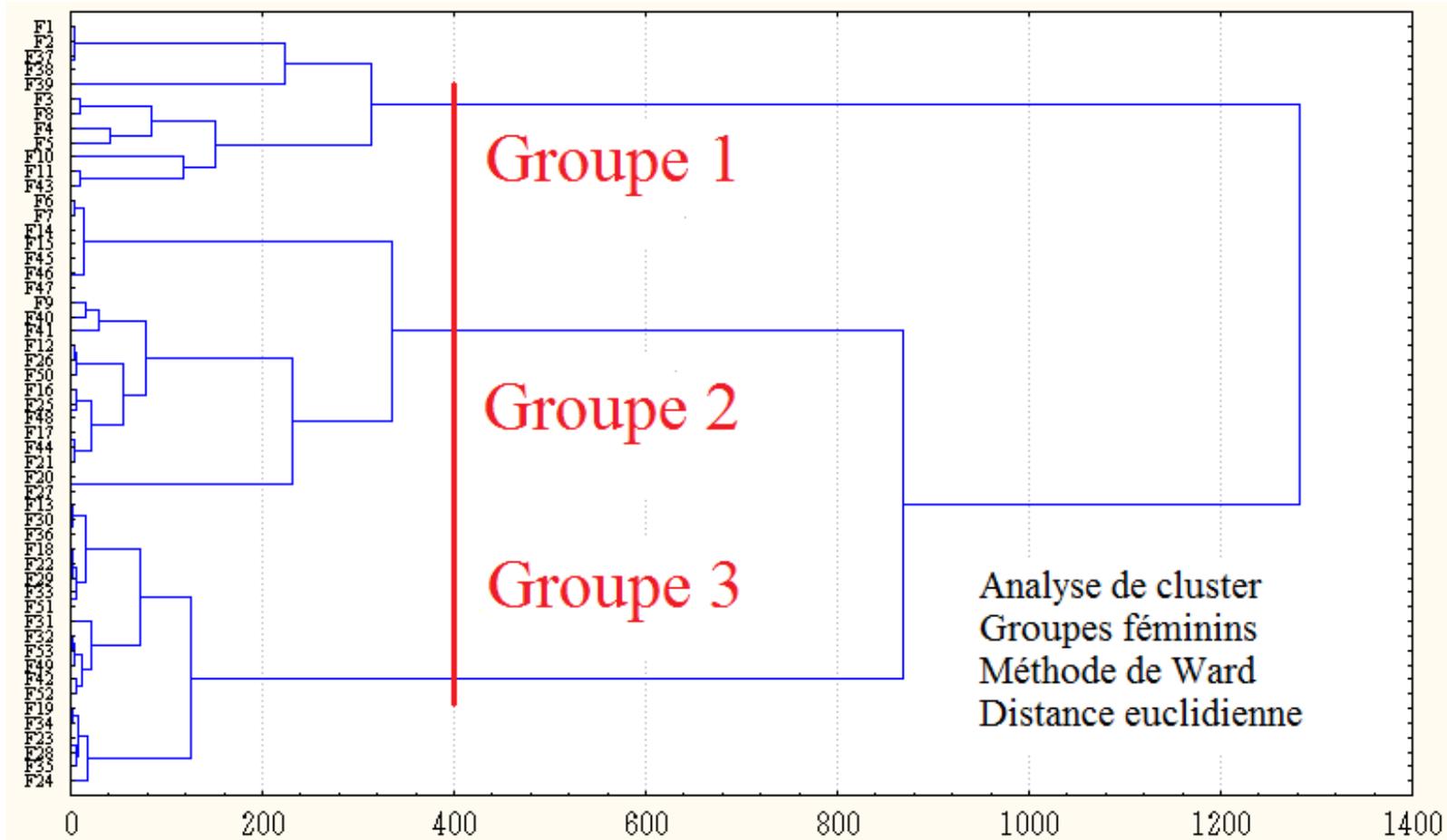
Variation chez les locuteurs

personnage	sexe	-nou	-naï	-née		-aï	-ée
Kaki no itchi	M	4 (100%)	0 (0%)	0 (0%)		7 (100%)	0 (0%)
Fourouzaem on	M	7 (77.8%)	1 (11.1)	1 (11.1%)		31 (91.2%)	3 (8.8%)
Katsoubée	M	0 (0%)	0 (0%)	16 (100%)		3 (6.4%)	44 (93.6%)
. . . .							
Oyasou	F	3 (100%)	0 (0%)	0 (0%)		42 (98%)	1 (2%)
Ohanée	F	0 (0%)	0 (0%)	2 (100%)		7 (41.2%)	10 (58.8%)
Osarou	F	1 (11.1%)	0 (0%)	8 (88.9%)		6 (18.8%)	26 (81.3%)

Analyse de cluster (45 personnages masculins)



Analyse de cluster (53 personnages féminins)



Calcul de pourcentages des variantes

Analyse sociolinguistique de la variation

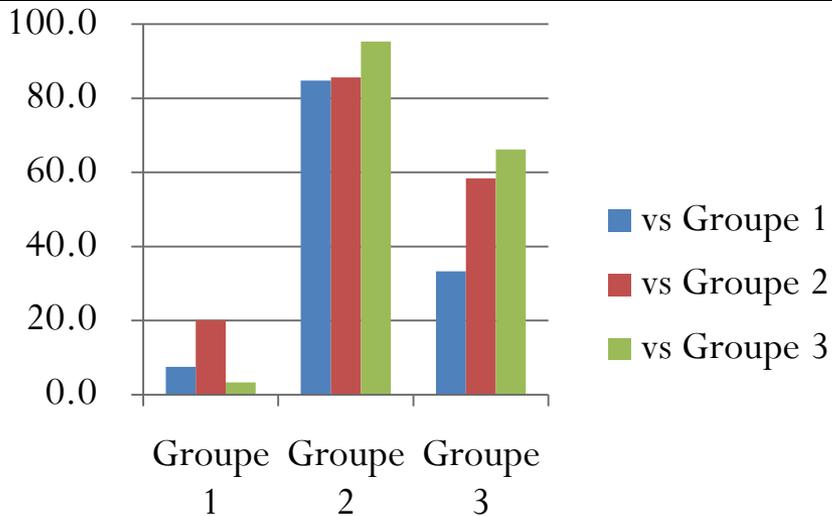
hommes	-nou	-nai	-née	-ai	-ée
Groupe 1	90.4	1.6	8.0	94.5	5.5
Groupe 2	9.8	0.0	90.2	56.9	30.6
Groupe 3	1.1	0.0	98.9	12.8	87.2

femmes	-nou	-nai	-née	-ai	-ée
Groupe 1	59.7	21.4	18.9	89.7	10.3
Groupe 2	2.6	0.0	97.4	71.9	18.6
Groupe 3	0.8	0.0	99.2	31.7	68.3

	Hommes	Femmes	Couche sociale	Variantes
Groupe 1	INKYO “Retraité”, ISHA “docteur”, HAÏKAÏSHI “poète de <i>haïkou</i> ”, KIKAKOU dit le Maître, NAMAYOÏ “samouraï”	OYASOU et OTATSOU “épouses de commerçants”	Haute ↑	-nou et -ai prédominantes
Groupe 2	MOUDASOUKÉ et TOBIHATCHI “hommes débauchés”, BANTOO “commis”	OKAKO et OSAMI “geïshas”	Non-haute / Non-basse	-née et -ai prédominantes
Groupe 3	YOUKOU MI “serveur de bain”, ISAMI et TEPPOSAKOU “hommes débauchés”	OMAROU bonne”, OSHITA et OTOBI “filles publiques”, OTAÏ “serveuse”	↓ Basse	-née et -ée prédominantes

Comportements linguistiques différents

groupe	locuteur	groupe	interlocuteur	-aï	-ée:	taux de variante -ée
1	OTATSOU	1	OOUMA	3	3	0.50
1	OTATSOU	1	TATSOUMI	31	0	0.00
1	INKYO	2	BANTOO	8	10	0.56
1	MATSOUÉMON	3	HACHIBEE	29	1	0.03
2	MOUDASOUKÉ	1	KANJI	15	18	0.55
2	OHANÉ	2	BABAMOJI	2	2	0.50
2	OHANÉ	2	TOYONEKO	0	3	1.00
2	OHANÉ	3	OKAKO	1	4	0.80
3	TOBIHACHI	1	KIKAKOU	4	2	0.33
3	OSAMI	2	OTAÏ	4	0	0.00



cf. Nakata (1985)

- Emploi très faible de *-ée* dans Groupe 1
- Emploi fréquent de *-ée* dans Groupe 2
- Code-switching dans Groupe 3

Pour conclure

- Problèmes de la caractérisation du français parlé
- Caractère contradictoire de la langue parlée dans les textes écrits
- Nécessité de l'analyse variationnelle de la langue parlée dans les textes
- Possibilité de surmonter les difficultés de l'analyse de la variation dans les textes écrits

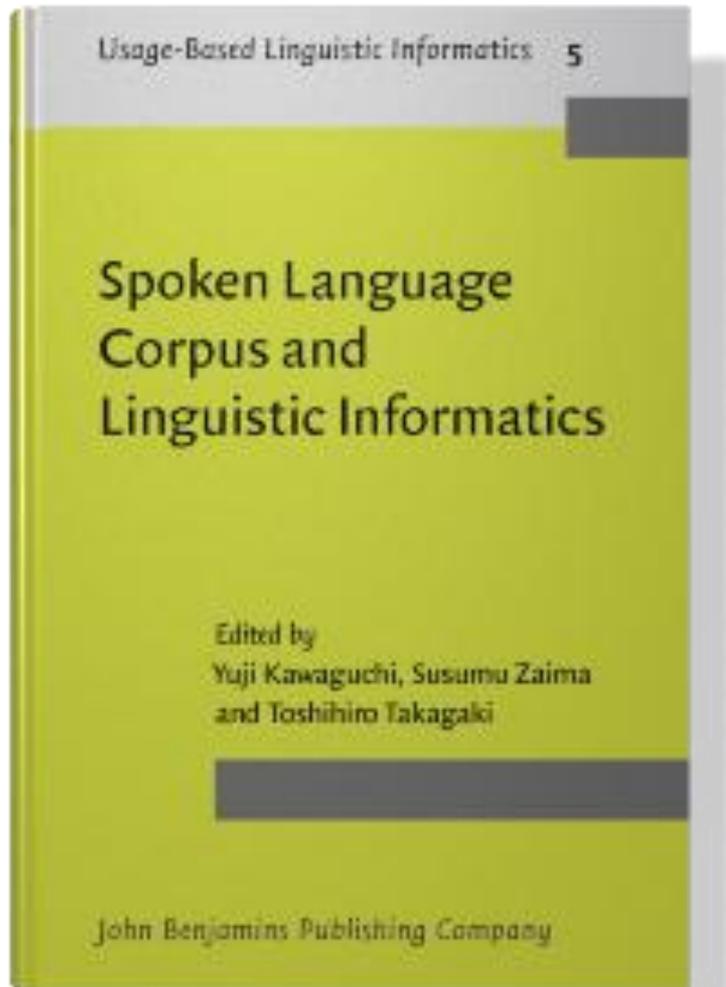
Références

- Ayres-Bennett Wendy (2004) *Sociolinguistic Variation in Seventeenth-Century France*, London: Cambridge University Press.
- Blanche-Benveniste C. et Jeanjean C. (1987) *Le français parlé : transcription et édition*, Paris: Didier Erudition.
- Bougy Catherine (1996) "Une Mazarinade méconnue Dialogue de deux paysans normands au XVIIe siècle", *Histoire et sociétés rurales* 6, pp.43-164.
- Ernst Gerhard (1985) *Gesprochenes Französisch zu Beginn des 17. Jahrhunderts*, Tübingen: Max Niemeyer Verlag.
- Deulofeu Henri-José (2008) "Écritures anciennes, paroles actuelles: nouveaux regards sur le système, la norme et les usages", *Discours, Diachronie, Stylistique Du Français : Etudes En Hommage à Bernard Combettes*, O. Bertrand, S. Prevost et M. Charolles (eds.), Peter Lang, 2008, pp.193-205.
- De Gorog, R. (1990) "Early Seventeenth-Century Spoken French and Héroard's Journal", *Romance Philology* 43, pp.431-442.
- Foisil Madeleine (éd.) (1989) *Journal de Jean Héroard, médecin de Louis XIII*, 2 vol., Paris: Fayard.
- Fukushima Naoyasu (2002) *Lorsque <ABUNai> devient <ABUNee>*, Tokyo: Kasama Sho'in.
- Jinbô Kazuya (éd) (1989) *Oukiyoubouro*, Tokyo: Iwanami Shoten.
- Kawaguchi Yuji (1997) "Sur les signes conventionnels motivés du japonais", *Polyphonie pour Ivan Fónagy*, Paris: L'Harmattan, pp.205-213.
- Kawaguchi Yuji (2006) "Usage-Based Approach to Linguistic Variation -Evidence from French and Turkish-," *Spoken Language Corpus and Linguistic Informatics*, Y. Kawaguchi et al., John Benjamins, pp. 247-267.
- Kawaguchi Yuji, Makoto Minegishi, Jacques Durand (2009) *Corpus Analysis and Variation in Linguistics*, John Benjamins, 399p.
- Kristol Andres (2009) "Sociolinguistique historique et analyse de la conversation: une nouvelle approche du *Journal d'hygiène* de Jean Héroard", *Vox Romanica*, 68, 169-186.
- Koch Peter (1990) *Wulf Oesterreicher, Gesprochene Sprache in der Romania : Französisch, Italienisch, Spanisch*, Tübingen: Max Niemeyer.
- Komatsu Hisao (1985) *Langue japonaise à l'époque d'Edo*, Tokyo: Tokyôdô.
- Matsumura Akira (1977) *Langue nationale de l'époque moderne, de Edo jusqu'à maintenant*, (en jp.), Tokyo: Ôfûsha.
- Matsumura Akira (1998) *Recherches sur les langues de Edo et de Tokyo*, nouvelle édition (en jp.), Tokyo: Tokyôdô.
- 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'allongement vocalique de *-ai-*", (en jp) *Jinbun gaku* 173, pp.25-47.



Atelier coopératif de C-ORAL-ROM et UBLI:
Corpus de langue parlée – son importance et
application, septembre 2005, à Tokyo University of
Foreign Studies

Merci de votre attention !



Spoken Language Corpus and Linguistic Informatics

Edited by Yuji Kawaguchi, Susumu
Zaima and Toshihiro Takagaki Tokyo
University of Foreign Studies

2006, John Benjamins,
Amsterdam/Philadelphia