

ÉTUDE DE CAS BASÉE SUR UNE TÂCHE DE « CONVERSATION LIBRE » DANS LE CADRE DE L'IPFC

-ANALYSE DE RICHESSE LEXICALE DU FRANÇAIS
CHEZ DES APPRENANTS JAPONAIS -

Kaori SUGIYAMA

Doctorante à l'Université Bordeaux 3 (CLLE-ERSSàB) et
à l'Université des Langues Étrangères de Tokyo

sugiyamakaori@gmail.com

Grandes lignes



0. Introduction
1. Littérature
2. Objectifs et postulats
3. Question de recherche
4. Données
5. Méthode
6. Résultat
7. Conclusions et Discussions

O. introduction: questions de départ

- IPFC → Analyse de l'interPHONOLOGIE chez les apprenants
- Corpus de l'IPFC → données sur la production orale en français langues étrangères
= exploitable pour des analyses outre la phonologie
- Étude lexicale en se basant sur le corpus de l'IPFC-Japonais

O. introduction: question de départ

- Augmentation du nombre des études de l'utilisation lexicale sur la production écrite chez les apprenants



- Peu d'études sur la production orale
- Aucune étude sur l'usage de mots par les apprenants Japonais du français dans la production orale

1. Littérature

richesse lexicale basée sur le corpus

- Développé dans le domaine des études sur la production écrite
 - Richesse lexicale → concept multi-facette (Tidball and Treffers-Daller 2007:134)
 - diversité
sophistication
densité
nb d'erreurs
- (Read 2000: 200-201)

1. Littérature

richesse lexicale basée sur le corpus

- Développé dans le domaine des études sur la production écrite
- Richesse lexicale → concept multi-facette (Tidball and Treffers-Daller 2007:134)

- diversité
sophistication
densité
nb d'erreurs

(Read 2000: 200-201)

1. Littérature diversité

- Variété des différents mots plutôt qu'un nb limité de mots répétés (Read 2000: 200)
-

1. Littérature diversité

- Variété des différents mots plutôt qu'un nb limité de mots répétés (Read 2000: 200)

Nb de mots différents
employés

Nb total de mots
employés

- Measure \rightarrow TTR = **type/token**

1. Littérature diversité

- Variété des différents mots plutôt qu'un nb limité de mots répétés (Read 2000: 200)

- Measure \rightarrow TTR = **type/token**



transformation

(Bulté et al. 2008: 279)

Indice de Guiraud = **type/** $\sqrt{\text{token}}$

\rightarrow statistiquement correct

(Van Hout and Vermeer 2007:100-102, Treffers-Daller 2009:82)

1. Littérature diversité

problèmes

- Aucune considération pour la fréquence d'un mot (Van Hout and Vermeer 2007:106)
- Tous les items lexicaux comme valeur égale
- =nécessaire d'être compensé par la dimension **qualitative** (Malvern and Richards 2009:165)

1. Littérature diversité

problèmes

- Aucune considération pour la fréquence d'un mot (Van Hout and Vermeer 2007:106)
- Tous les items lexicaux comme valeur égale
- =nécessaire d'être compensé par la dimension **qualitative** (Malvern and Richards 2009:165)



Sophistication

aspect qualitatif de la compétence productive chez les apprenants (Bulté et al. 2008: 286)

1. Littérature sophistication

□ Lexical Frequency Profile (LFP)

(Laufer and Nation 1995:311-313)

→ montre le pourcentage de mots que les apprenants utilisent dans les différentes zones de fréquence du vocabulaire

→ basé sur les “**word families**”

1. Littérature sophistication

□ Lexical Frequency Profile (LFP)

(Laufer and Nation 1995:311-313)

→ montre le pourcentage de mots que les apprenants utilisent dans les différentes zones de fréquence du vocabulaire

→ basé sur les “**word families**”

Unité plus grande que lemme = forme canonique (même radicale)

□ Unité contenant les formes dérivées (affixe)

1. Littérature

VocabProfil

- Version française de LFP = **vocabprofil***
 - basé sur le corpus des journaux (*Le Monde* et *Le Soir*) de 50 million de mots publiés en 1998
(Verlinde and Selva 2001)

1. Littérature

VocabProfil

□ Version française de LFP = **vocabprofil***

→ basé sur le corpus des journaux (*Le Monde* et *Le Soir*) de 50 million de mots publiés en 1998

(Verlinde and Selva 2001)

→ 4 zones de fréquence

1 ~ 1000 =K1 (**mots de haute fréq.**)

1001 ~ 2000 =K2 (**mots de fréq. moyenne**)

2001 ~ 3000 =K3 (**mots de basse fréq.**)

des autres =MHL (**mots hors liste**)

1. Littérature

VocabProfil

□ Version française de LFP = **vocabprofil***

→ basé sur le corpus des journaux (*Le Monde* and *Le Soir*) de 50 million de mots publiés en 1998

(Verlinde and Selva 2001)

→ 4 zones de fréquence

1 ~ 1000 =K1 (**mots de haute fréq.**)

1001 ~ 2000 =K2 (**mots de fréq. moyenne**)

2001 ~ 3000 =K3 (**mots de basse fréq.**)

des autres =MHL (**mots hors liste**)

→ calcul automatique de “type”, “token”, “famille de mots” de chaque zone de fréquence

1. Littérature

VocabProfil et production orale

□ Ovtcharov et al.(2006)

comparaison des ratios de chaque zone de fréq.

*niv. intermédiaire ↔ avancé

différence significative → fréq. moyenne & basse

*avancé ↔ NNS

pas de différence significative

*ratio de basse fréq. ↔ niveaux

forte corrélation

1. Littérature

VocabProfil et production orale

□ Ovtcharov et al.(2006)

comparaison des ratios de chaque zone de fréq.

*niv. intermédiaire ↔ avancé

différence significative → fréq. moyenne & basse

*avancé ↔ NNS

pas de différence significative

*ratio de basse fréq. ↔ niveaux

forte corrélation

1. Littérature

VocabProfil et production orale

□ Thomas (2008)

Progrès lexicale après l'expérience à l'étrangère

*groupe expérimenté vs groupe contrôlé

Comparaison de ratio de chaque zone de fréq.

pré-test ⇔ post-test

aucune différence significative dans toutes les zones

→ token ⇒ 11.7% ↑ vs 1.7% ↑

→ différence qualitative

(non -standard ↑ : abréviations, expressions familières)

2. Objectifs et postulats

objectifs

- Investigation de richesse lexicale chez les apprenants Japonais de français par l'analyse comparative (apprenants vs locuteurs natifs, apprenants vs apprenants)

postulats

- Les mots de haute fréq. sont plus faciles à acquérir.
- La production des mots de basse fréq. est plus difficile

3. Questions de recherche

1. Les locuteurs natifs ont-ils un indice de richesse lexicale plus haut que les apprenants?
2. La richesse lexicale peut-il prédire la compétence linguistique?
3. Les apprenants de niveau avancé ont-ils un profil plus proche des locuteurs natifs?
4. Chaque groupe classifié selon la richesse lexicale a-t-il une différente sorte de vocabulaire?

4. Données

□ Participants

28 étudiants Japonais de français avec au moins 3 ans d'études en français

→ conversation libre dans le corpus du projet IPFC

14 locuteurs natifs

→ conversation libre dans le corpus d'Aix-Marseille de TUFs*¹

□ Élaboration des données

→ Élimination de certains items (non propres, marqueurs d'hésitation, mots d'origine japonaise etc...)

=focus strict sur le lexique français indépendant du choix de topique.

	type	token
Corpus d'apprenants	3864	12276
Corpus de locuteurs natifs	8725	44711

*¹http://www.coelang.tufs.ac.jp/multilingual_corpus/fr/index.html?contents_xml=top&menulang=en

5. Méthode

participants

	Haut M-outils	Haut M-pleins	Moyen	Bas	Hors-liste
AP_1	2.56	5.90	2.85	1.15	2.98
AP_2	3.07	7.95	4.04	2.67	4.31
AP_3	2.54	8.01	3.83	1.81	2.92
AP_4	2.75	6.69	2.77	0.58	3.25
AP_5	2.80	7.56	2.79	1.79	4.35
AP_6	2.48	6.67	3.21	1.73	1.51
AP_7	2.73	6.11	1.51	1.15	1.22

5. Méthode

Zone de fréq.

Mots-outils de haute fréq

→ 1 ~ 1000 (function words)

Mots-pleins de haute fréq.

→ 1 ~ 1000 (content words)

fréq. moyenne

→ 1001 ~ 2000

Basse fréq.

→ 2001 ~ 3000

Hors liste

→ 3001 ~

	Haut M-outils	Haut M-pleins	Moyen	Bas	Hors-liste
AP_1	2.56	5.90	2.85	1.15	2.98
AP_2	3.07	7.95	4.04	2.67	4.31
AP_3	2.54	8.01	3.83	1.81	2.92
AP_4	2.75	6.69	2.77	0.58	3.25
AP_5	2.80	7.56	2.79	1.79	4.35
AP_6	2.48	6.67	3.21	1.73	1.51
AP_7	2.73	6.11	1.51	1.15	1.22

5. Méthode

Indice de Guiraud

Ex: $\text{type de K2} / \sqrt{\text{token de K2}}$

	Haut M-outils	Haut M-pleins	Moyen	Bas	Hors-liste
AP_1	2.56	5.90	2.85	1.15	2.98
AP_2	3.07	7.95	4.04	2.67	4.31
AP_3	2.54	8.01	3.83	1.81	2.92
AP_4	2.75	6.69	2.77	0.58	3.25
AP_5	2.80	7.56	2.79	1.79	4.35
AP_6	2.48	6.67	3.21	1.73	1.51
AP_7	2.73	6.11	1.51	1.15	1.22

6. Résultat

□ Statistique descriptive

	Moyenne		Valeur minimum		Valeur maximum		Écart type	
	LN	AP	LN	AP	LN	AP	LN	AP
Haut M-outils	2.43	3.00	1.79	2.28	2.95	3.81	0.37	0.41
Haut M-pleins	11.10	6.21	8.32	3.90	17.19	8.87	2.23	1.29
Moyen	6.63	2.59	3.93	1.34	10.80	4.04	2.05	0.73
Bas	4.01	1.35	1.63	0.00	7.63	3.00	1.53	0.75
Hors-liste	7.98	2.38	4.12	1.22	13.02	4.13	2.52	0.78

6. Résultat

□ Statistique descriptive

	Moyenne		Valeur minimum		Valeur maximum		Écart type	
	LN	AP	LN	AP	LN	AP	LN	AP
Haut M-outils	2.43	3.00	1.79	2.28	2.95	3.81	0.37	0.41
Haut M-pleins	11.10	6.21	8.32	3.90	17.19	8.87	2.23	1.29
Moyen	6.63	2.59	3.93	1.34	10.80	4.04	2.05	0.73
Bas	4.01	1.35	1.63	0.00	7.63	3.00	1.53	0.75
Hors-liste	7.98	2.38	4.12	1.22	13.02	4.13	2.52	0.78

6. Résultat

□ Corrélation

	Haut M-outils	Haut M-pleins	Moyen	Bas	Hors-liste
Haut M-outils	1				
Haut M-pleins	-0.50	1			
Moyen	-0.67	0.78	1		
Bas	-0.60	0.69	0.92	1	
Hors-liste	-0.66	0.77	0.94	0.90	1

6. Résultat

□ Corrélation

	Haut M-outils	Haut M-pleins	Moyen	Bas	Hors-liste
Haut M-outils	1				
Haut M-pleins	-0.50	1			
Moyen	-0.67	0.78	1		
Bas	-0.60	0.69	0.92	1	
Hors-liste	-0.66	0.77	0.94	0.90	1

6. Résultat

- Analyse en Composante Principales

	Valeur propre	Taux de compression %	Valeur propre cumulée	% cumulé de variance
1^{ère} composante principale	4.00	80.07	4.00	80.07
2^{ème} composante principale	0.53	10.57	4.53	90.64

6. Résultat

- Analyse en Composante Principales

	Valeur propre	Taux de compression %	Valeur propre cumulée	% cumulé de variance
1^{ère} composante principale	4.00	80.07	4.00	80.07
2^{ème} composante principale	0.53	10.57	4.53	90.64

6. Résultat

- Poids factoriels

	1 ^{ère} composante principale
Haut M-outils	-0.75
Haut M-pleins	0.84
Moyen	0.97
Bas	0.93
Hors-liste	0.96

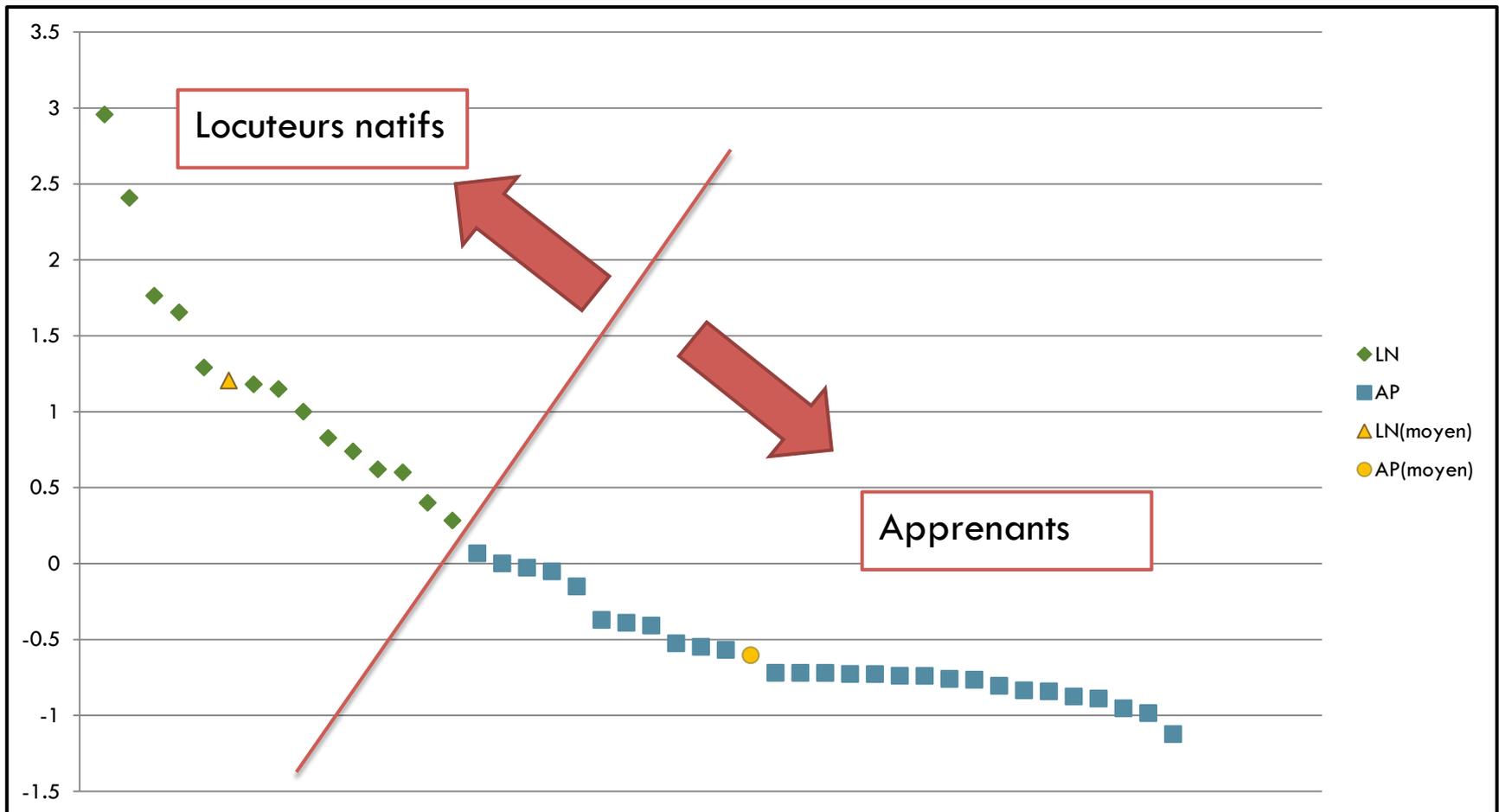
6. Résultat

- Poids factoriels

	1 ^{ère} composante principale
Haut M-outils	-0.75
Haut M-pleins	0.84
Moyen	0.97
Bas	0.93
Hors-liste	0.96

6. Résultat

• Score individuel



6. Résultat

□ Groupes

	Doctorat	Master	Licence	Spécialité en linguistique française	Expérience dans des pays francophones
A	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>
B		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
C			<input type="radio"/>		

6. Résultat

	LN		A		B		C
47	truc	9	fac	7	galettes	10	film
35	super	6	ok	6	nager	7	bain
24	fac	5	merci	3	alcool	7	fondue
23	prof	4	galette	3	merci	7	pomme
21	trucs	3	chocolat			5	super
19	ok	3	franchement			4	candide
18	profs	3	frites			4	dvd
15	comparée	3	gentils			4	empaillages
15	foi					4	voyagé
14	franchement	3	rédigé			3	bizarre
14	semestre	3	super			3	ok
13	filières					3	sauce
12	bac						
12	sympa						
11	boulot						
11	tandem						
11	élus						
10	bla						
10	merci						

Troncations

6. Résultat

	LN		A		B		C
47	truc	9	fac	7	galettes	10	film
35	super	6	ok	6	nager	7	bain
24	fac	5	merci	3	alcool	7	fondue
23	prof	4	galette	3	merci	7	pomme
21	trucs	3	chocolat			5	super
19	ok	3	franchement			4	candide
18	profs					4	dvd
15	comparée	3	gentils			4	empaillages
15	foi	3	noël			4	voyagé
14	franchement	3	rédigé			3	bizarre
14	semestre	3	super			3	ok
13	filières					3	sauce
12	bac						
12	sympa						
11	boulot						
11	tandem						
11	élus						
10	bla						
10	merci						

Mots familiers

6. Résultat

	LN		A		B		C		
47	truc		9	fac		7	galettes	10	film
35	super					6	nager	7	bain
24	fac					3	alcool	7	fondue
23	prof		4	galette		3	merci	7	pomme
21	trucs		3	chocolat				5	super
19	ok		3	franchement				4	candide
18	profs		3	frites				4	dvd
15	comparée		3	gentils				4	empaillages
15	foi		3	noël				4	voyagé
14	franchement		3	rédigé				3	bizarre
14	semestre		3	super				3	ok
13	filières							3	sauce
12	bac								
12	sympa								
11	boulot								
11	tandem								
11	élus								
10	bla								
10	merci								

remplisseur

6. Résultat

	LN		A		B		C
47	truc	9	fac	7	galettes	10	film
35	super	6	ok	6	nager	7	bain
24	fac	5	merci	3	alcool	7	fondue
23	prof	4	galette	3	merci	7	pomme
21	trucs	3	chocolat			5	super
19	ok	3	franchement			4	candide
18	profs	3	frites			4	dvd
15	comparée	3	gentils			4	empaillages
15	foi	3	noël			4	voyagé
14	franchement	3	rédigé			3	bizarre
14	semestre	3	super			3	ok
13	filières					3	sauce
12	bac						
12	sympa						
11	boulot						
11	tandem						
11	élus						
10	bla						
10	merci						

7. Conclusions et Discussions

- Les locuteurs natifs ont-ils l'indice de richesse lexicale plus haute que les apprenants?

 oui

- La richesse lexicale peut-il prédire la compétence linguistique?

 oui

7. Conclusions et Discussions

- Les apprenants de niveau avancé ont-ils le profil plus proche de celui de locuteurs natifs?



oui

- Chaque groupe classifié selon la richesse lexicale a-t-il une différente sorte de vocabulaire?



oui

7. Conclusions et Discussions

- Grande différence du score parmi les locuteurs natifs
- Autres niveaux que les lexiques
ex. collocation, différence de l'utilisation du temps
- Contribution au domaine du FLE
quels mots à enseigner? Les mots de basse fréq.?
fréq. de mots dans la production orale?
- Mots particuliers à l'oral
quand enseigner ces mots?

bibliographie

- Bulté, B., A. Housen, M. Pierrard and S. van Daele. 2008. “Investigating Lexical Proficiency Development over Time—The case of Dutchspeaking Learners of French in Brussels”. *French Language Studies* 18. 277-298.
- Laufer, B. and P. Nation. 1995. “Vocabulary Size and Use: Lexical richness in L2 written production”. *Applied Linguistics* 16(3). 307-322.
- Malvern, D. and B. Richards. 2009. “A New Method of Measuring Rare Word Diversity: The Example of L2 learners of French”. *Vocabulary Studies in First and Second Language Acquisition—The Interface between Theory and Application*, Richard, B., H. Daller., D. Malvern., P. Meara., J. Milton and J. Treffers-Daller (eds). Basingstoke: Palgrave Macmillan. 164-178.
- Ovtcharov, V., T. Cobb and R. Halter. 2006. “La Richesse Lexicale des Productions Orales: Mesure Fiable du Niveau de Compétence Langagière”. *The Canadian Modern Language Review* 63(1). 107-125.
- Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press

bibliographie

- Thomas, A. 2008. “La Mesure des Progrès Lexicaux en FL2 Avancé”. *Congrès Mondial de Linguistique Française—CMLF08*. Institut de Linguistique Française, J. Durand, B. Habert and B. Laks (eds). 587-597.
- Tidball, F. and J. Treffers-Daller. 2007. “Exploring Measures of Vocabulary Richness in Semi-Spontaneous French Speech—A quest for the Holy Grail?”. *Modelling and Assessing Vocabulary Knowledge*, Daller, J., J. Milton and J. Treffers-Daller (eds). Cambridge: Cambridge University Press. 133-149.
- Treffers-Daller, J. 2009. “Language Dominance and Lexical Diversity: How bilinguals and L2 learners differ in their knowledge and use of French lexical and functional items”. *Vocabulary Studies in First and Second Language Acquisition—The Interface between Theory and Application*, Richard, B., H. Daller., D. Malvern., P. Meara, J. Milton and J. Treffers-Daller (eds). Basingstoke: Palgrave Macmillan. 74-90.
- van Hout, R. and Vermeer, A. 2007. “Comparing Measures of Lexical Richness”. *Modelling and Assessing Vocabulary Knowledge*, Daller, J., J. Milton and J. Treffers-Daller (eds). Cambridge: Cambridge University Press. 93-114.
- Verlinde, S. and Selva, T. 2001. “Corpus-Based Versus Intuition-Based Lexicography: Defining a word list for a French learner’s dictionary”. *Proceedings of the Corpus Linguistics 2001 conference. Technical Papers 13*, P. Rayson, A., T. Wilson., A. McEnery and S. Khoja (dir).

MERCI POUR VOTRE ATTENTION

Cette recherche est soutenue par le Global COE Program,
« Corpus-based Linguistics and Language Education »
(CbLLE) de l'Université des Langues Etrangères de Tokyo