Lexical Profile of French Learner Speech

- in case of Japanese University Students-

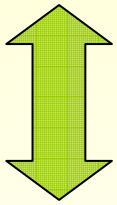
Kaori SUGIYAMA
Graduate student
of Tokyo University of Foreign Studies
Yuji KAWAGUCHI
Tokyo University of Foreign Studies

Outline

0. Introduction
1. Literature review
2. Objectives and RQ
3. Method
4. Result
5. Discussion and Conclusion

0. Introduction :setting

- Increasing number of learners' vocabulary in use of English and other languages



- Few studies of learners spontaneous speech

- No studies about vocabulary use in speech in case of Japanese students of French

1.Literature review

Lexical richness based on corpus analysis

- Developed in writing research domain

1.Literature review

Lexical richness based on corpus analysis

- Developed in writing research domain
- **Lexical richness**

→multi-faceted concept (Tidball and Treffers-Daller 2007:134)

- diversity
- sophistication
- complexity
- productivity
- fluency

(Bulté et al. 2008: 279)

1.Literature review

Lexical richness based on corpus analysis

- Developed in writing research domain
- **Lexical richness**

→multi-faceted concept (Tidball and Treffers-Daller 2007:134)

- **diversity**
- **sophistication**
- complexity
- productivity
- fluency

(Bulté et al. 2008: 279)

1.Literature review

Diversity

- **Variety of different words** rather than a limited number of words used repetitively (Read 2000: 200)
- Measure → TTR = **type/token**

transformations thereof

(Bulté et al. 2008: 279)

Index of Guiraud = **type/ √token**

→accurate in most studies

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

1.Literature review

Diversity

- **Variety of different words** rather than a limited [illegible] words u[illegible]vely (Read [illegible])
- Measure → TTR = **type/token**

Number of different words used

Total number of words used

transformations thereof

(Bulté et al. 2008: 279)

Index of Guiraud = **type/ √token**

→accurate in most studies

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

1.Literature review

Diversity

- **Variety of different words** rather than a limited number of words used repetitively (Read 2000: 200)
- Measure → TTR = **type/token**

transformations thereof

(Bulté et al. 2008: 279)

Index of Guiraud = **type/ √token**

→accurate in most studies

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

1.Literature review

Diversity

Problems

- No consideration for frequency of a word (Van Hout and Vermeer 2007:106)
- All vocabulary items as equal value

=need to be supplemented by **qualitative** dimension (Malvern and Richards 2009:165)

Sophistication

- **qualitative aspects** of learners' productive lexical proficiency (Bulté et al. 2008: 286)

1.Literature review

Diversity

Problems

- No consideration for frequency of a word (Van Hout and Vermeer 2007:106)
- All vocabulary items as equal value

=need to be supplemented by **qualitative** dimension (Malvern and Richards 2009:165)

Sophistication

- **qualitative aspects** of learners' productive lexical proficiency (Bulté et al. 2008: 286)

1.Literature review

Sophistication

- **Frequency-based** vocabulary measure (Bulté et al. 2008: 286)
- independent of syntax and text cohesiveness
  =**focus only on lexis** (Laufer and nation 1995:313)
- Narrower focus on specific or
  "advanced" words (Bulté et al. 2008: 280, Read 2000 :200)
  ||
- **low-frequency words** rather than just general, everyday vocabulary (Read 2000:200)

1.Literature review

Sophistication

- **Lexical Frequency Profile (LFP)**

(Laufer and nation 1995:311-313)

→show percentage of words a learner uses at different vocabulary frequency layers

→ based on "**word families**"

1.Literature review

Sophistication

- **Lexical Frequency Profile (LFP)**

(Laufer and nation 1995:311-313)

→show percentage of words a learner uses at different vocabulary frequency levels

→ based on "**word families**"

- Larger unit than lemma

= canonical form

(same stems and open classes)

- Unit also containing derived forms (affix)

(Ishikawa 2008: 78-81)

(ex: play, played, plays & player, replay etc…)

1.Literature review

VocabProfil

- **French version of LFP = vocabprofil***

  → based on newspaper corpus (*Le Monde* and *Le Soir*) of 50 million words published in 1998

  (Verlinde and Selva 2001)

→4 frequency layers

1～1000	=K1	(**High frequency words**)
1001～2000	=K2	(**middle frequency words**)
2001～3000	=K3	(**low frequency words**)
the others	=Off-List words	(**lower freq words**)

→automatic calculation of type, token, word family of each frequency layer

*http://www.lextutor.ca/vp/fr/

1.Literature review

VocabProfil

- **French version of LFP = vocabprofil***

  → based on newspaper corpus (*Le Monde* and *Le Soir*) of 50 million words published in 1998 (Verlinde and Selva 2001)

→4 frequency layers

1～1000 =K1 (**High frequency words**)
1001～2000 =K2 (**middle frequency words**)
2001～3000 =K3 (**low frequency words**)
the others =Off-List words (**not listed words**)

→automatic calculation of type, token, word family of each frequency layer

*http://www.lextutor.ca/vp/fr/

1.Literature review

VocabProfil

- **French version of LFP = vocabprofil***

→ based on newspaper corpus (*Le Monde* and *Le Soir*) of 50 million words published in 1998 (Verlinde and Selva 2001)

→4 frequency layers

1～1000 =K1 (**High freq words**)
1001～2000 =K2 (**middle freq words**)
2001～3000 =K3 (**low freq words**)
the others =Off-List words (**not listed words**)

→automatic calculation of type, token, word family of each frequency layer

*http://www.lextutor.ca/vp/fr/

1.Literature review

Vocabprofil and French learners' oral production

- Ovtcharov et al.(2006)

comparison of ratio of each freq layer

intermediate⇔advanced

middle & low freq

advanced⇔NNS

no significant difference

ratio of low freq ⇔levels

strong correlation

1.Literature review

Vocabprofil and French learners' oral production

- Thomas (2008)

lexical progress by studying abroad experience (experimental group vs control group)

Comparison of ratio of each freq layer

pretest ⇔posttest

no significant differences in all freq layers

(token ⇒ 11.7% ↑ vs 1.7% ↑)

qualitative differences

(non standard ↑ : abbreviations, slangs etc…)

2. Objectives and RQ

- Investigation of lexical richness in case of Japanese learners of French

previous studies ⇨ Sophistication
this study ⇨ Sophistication × diversity
(measure of diversity of rare words use
⇒more useful (Malvern and Richards 2009:165))

- What are differences of lexical richness among Japanese learners ?

3. Method

- **Participants**

  28 Japanese students with at least 3 years experience of learning French

  →non guided conversation corpus collected by the framework of IPFC Project*

  (Interphonologie du Français Contemporain)

- **Data elaboration**

  →elimination of certain items (Proper nouns, hesitation markers etc…)

*http://cblle.tufs.ac.jp/ipfc/

3. Method

- participants

	High all	High func	High cont	Middle	low	Off_list
AM	5.36	2.56	5.90	2.85	1.15	2.98
JB	7.11	3.07	7.95	4.04	2.67	4.31
KH	6.37	2.54	8.01	3.83	1.81	2.92
KK	6.15	2.75	6.69	2.77	0.58	3.25
KS	6.53	2.80	7.56	2.79	1.79	4.35
MH	5.77	2.48	6.67	3.21	1.73	1.51
MK	5.53	2.73	6.11	1.51	1.15	1.22

3. Method

□ **Frequency layers**

High all
→1～1000 (all)

High function words
→1～1000 (function words)

High content words
→1 ～1000(content words)

Middle
→1001～2000

Low
→2001～3000

Off list
→3001～

	High all	High func	High cont	Middle	low	Off_list
AM	5.36	2.56	5.90	2.85	1.15	2.98
JB	7.11	3.07	7.95	4.04	2.67	4.31
KH	6.37	2.54	8.01	3.83	1.81	2.92
KK	6.15	2.75	6.69	2.77	0.58	3.25
KS	6.53	2.80	7.56	2.79	1.79	4.35
MH	5.77	2.48	6.67	3.21	1.73	1.51
MK	5.53	2.73	6.11	1.51	1.15	1.22

3. Method

	High all	High func	High cont	Middle	low	Off list
AM	5.36	2.56	5.90	2.85	1.15	2.98
JB	7.11	3.07	7.95	4.04	2.67	4.31
KH	6.37	2.54	8.01	3.83	1.81	2.92
KK	6.15	2.75	6.69	2.77	0.58	3.25
KS	6.53	2.80	7.56	2.79	1.79	4.35
MH	5.77	2.48	6.67	3.21	1.73	1.51
MK	5.53	2.73	6.11	1.51	1.15	1.22

□ Index of Guiraud

ex: **type of K2/√token of K2**

3. Method

- Comparing averages of low freq (type, token, % of each freq layer)

general tendency of Japanese learners

- Cluster analysis

Classifying Japanese learners

- ANOVA

Test of significant differences among groups

- Correlation

Relationship among each freq layer

4. Result

Descriptive statistics

	TYPE	TOKEN
SUM	3839	11971
AVERAGE	137.11	427.54
SD	58.22	278.33

□ Great dispersion ⇒ NOT homogeneous

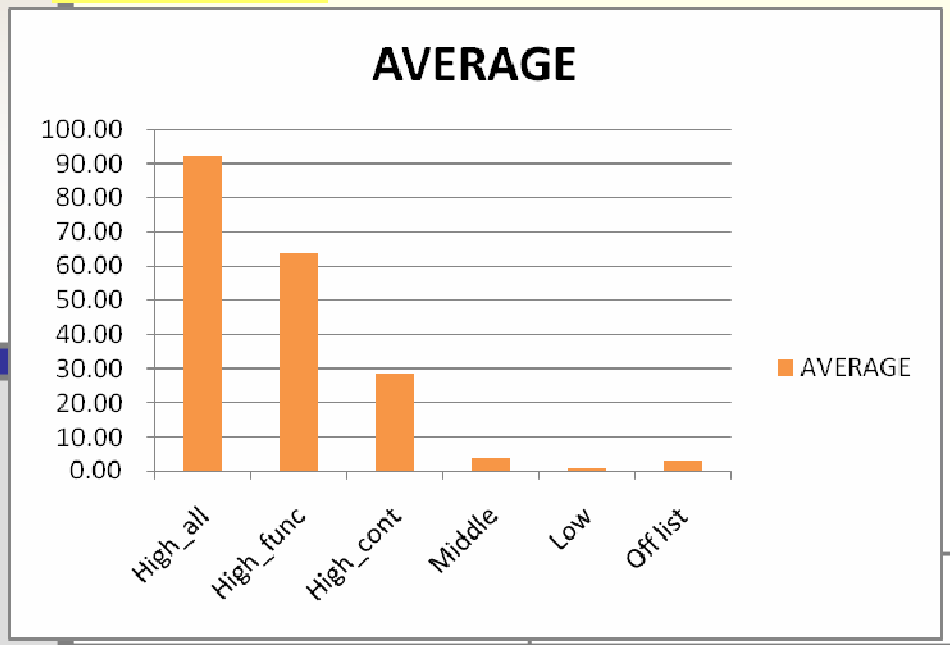
	3 years ⟷ 10 years+
No experience ↕ 3 years in French speaking countries	**sub groups?**

4. Result

Descriptive statistics

□ Ratio of each freq layers (Tokens)

	High_all	High_func	High_cont	Middle	Low	Off list
AVERAGE	92.33	63.62	28.72	3.56	1.07	3.00
SD	2.25	3.56	3.35	1.21	1.31	1.19

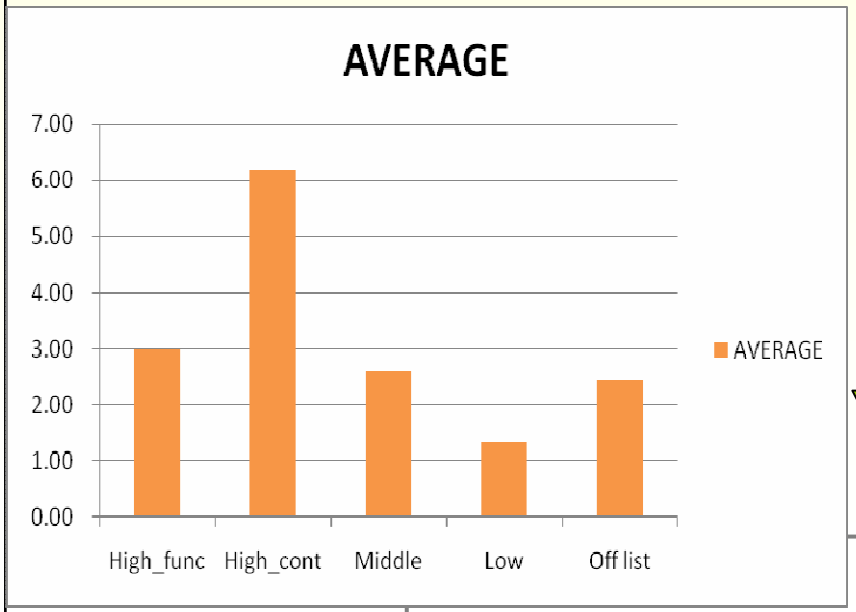


Great predominance by high freq words

(function words)

4. Result
Descriptive statistics
Index of Guiraud

	High func	High cont	Middle	low	Off list
AVERAGE	3.00	6.20	2.61	1.31	2.44
MAX	3.85	8.95	4.17	3.00	4.35
MIN	2.29	3.90	1.15	0.00	1.22
SD	0.43	1.30	0.75	0.74	0.83



High_content

High_function

Middle

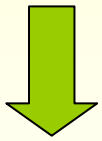
Off list

Low

4. Result

Descriptive statistics

- Low freq words vs Off list words
  (2001～3000)　　(3000～)
- Vocabprofil → based on journal corpus



Not including highly used words in daily use in conversation

(ex: super, OK, fac etc…)

4. Result

Cluster analysis

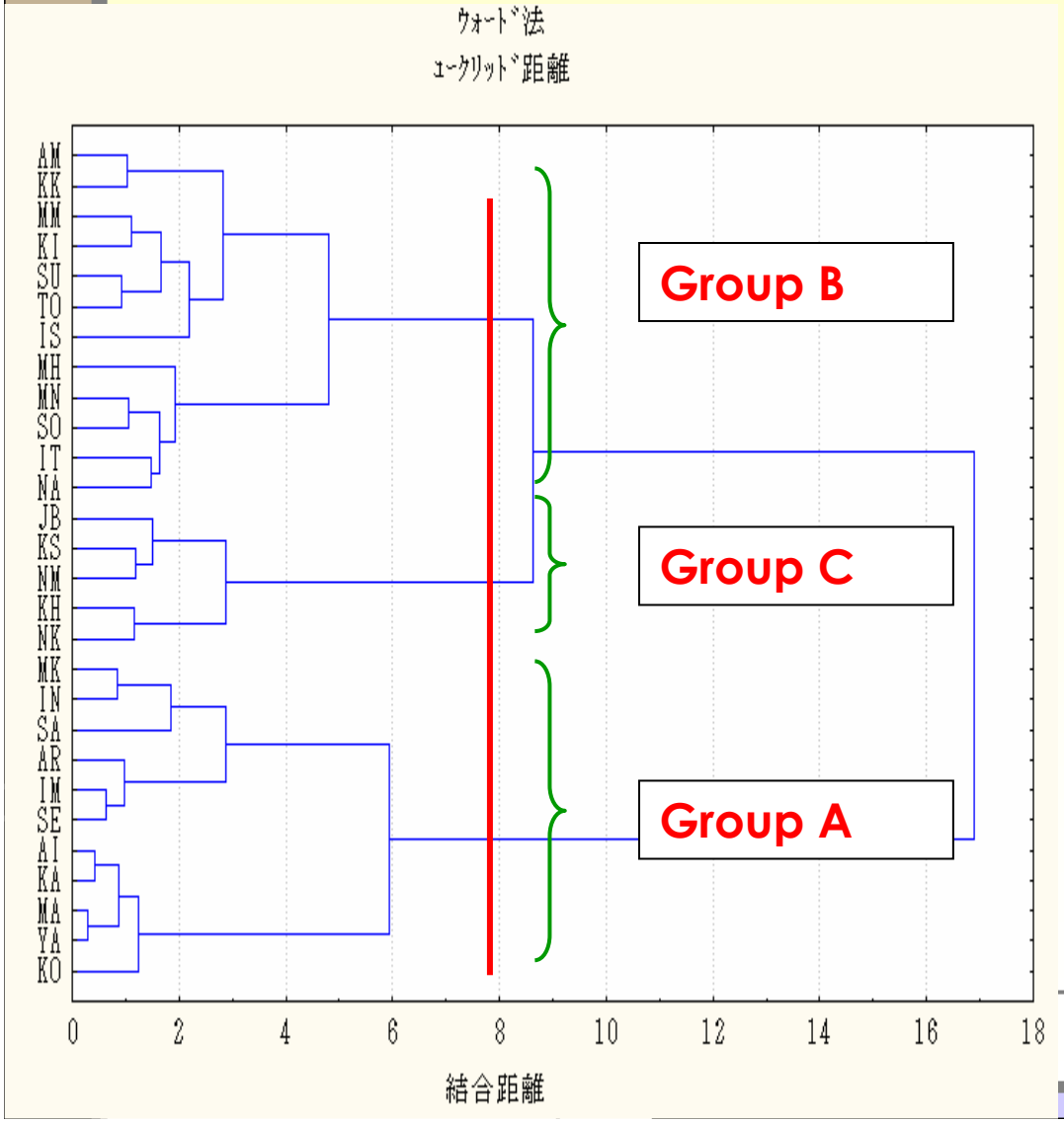
	TYPE	TOKEN
SUM	3839	11971
AVERAGE	137.11	427.54
SD	58.22	278.33

- Great dispersion ⇒ NOT homogeneous

	3 years ⟷ 10 years+
No experience ↕ 3 years in French speaking countries	**sub groups?**

4. Result

Cluster analysis



- Ward's method,
- Euclidean distance
- 3 groups

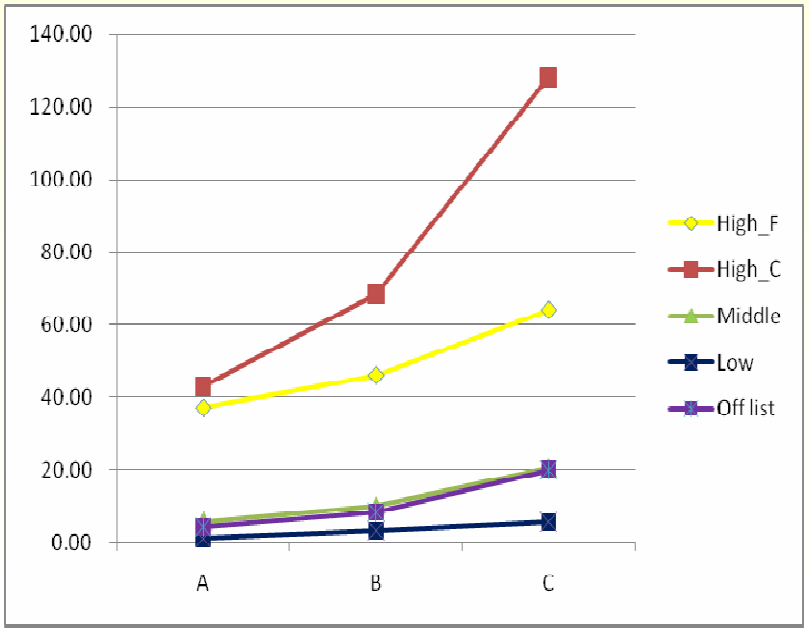
A→ 11

B→ 12

C→ 5

4. Result
Cluster analysis descriptive statistics

		High_F	High_C	Middle	Low	Off list
TYPE	A	37.09	42.91	5.82	1.09	4.36
	B	46.08	68.50	10.00	3.25	8.33
	C	64.20	128.20	20.60	5.60	20.00

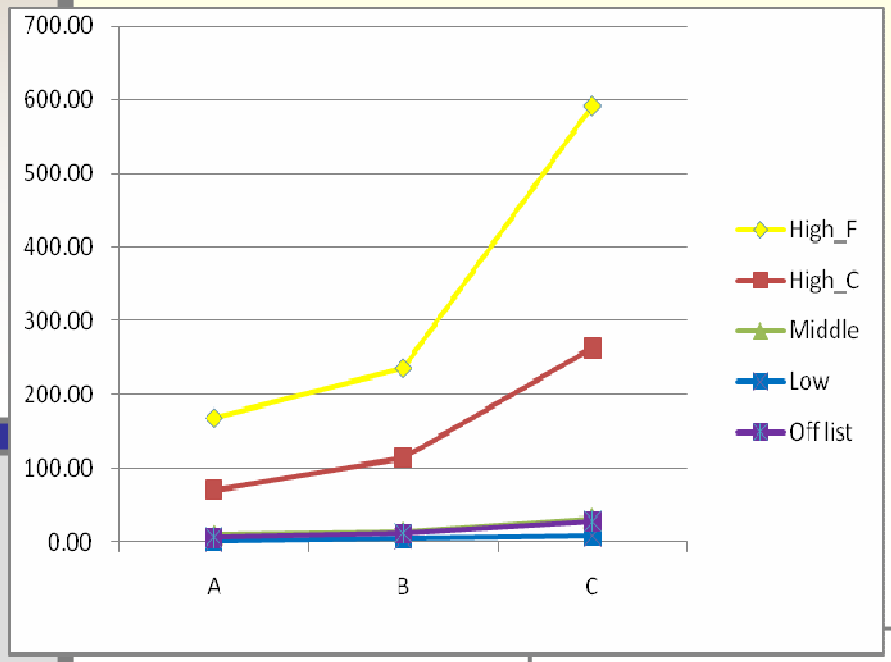


- Diversity → A < B < C
- Remarkable increase(High)

  B → C
- Few low freq word types

4. Result
Cluster analysis descriptive statistics

		High_F	High_C	Middle	Low	Off list
TOKEN	A	168.00	70.64	9.00	1.27	5.82
	B	235.58	114.33	13.58	4.25	11.25
	C	591.60	263.40	31.60	7.80	27.60

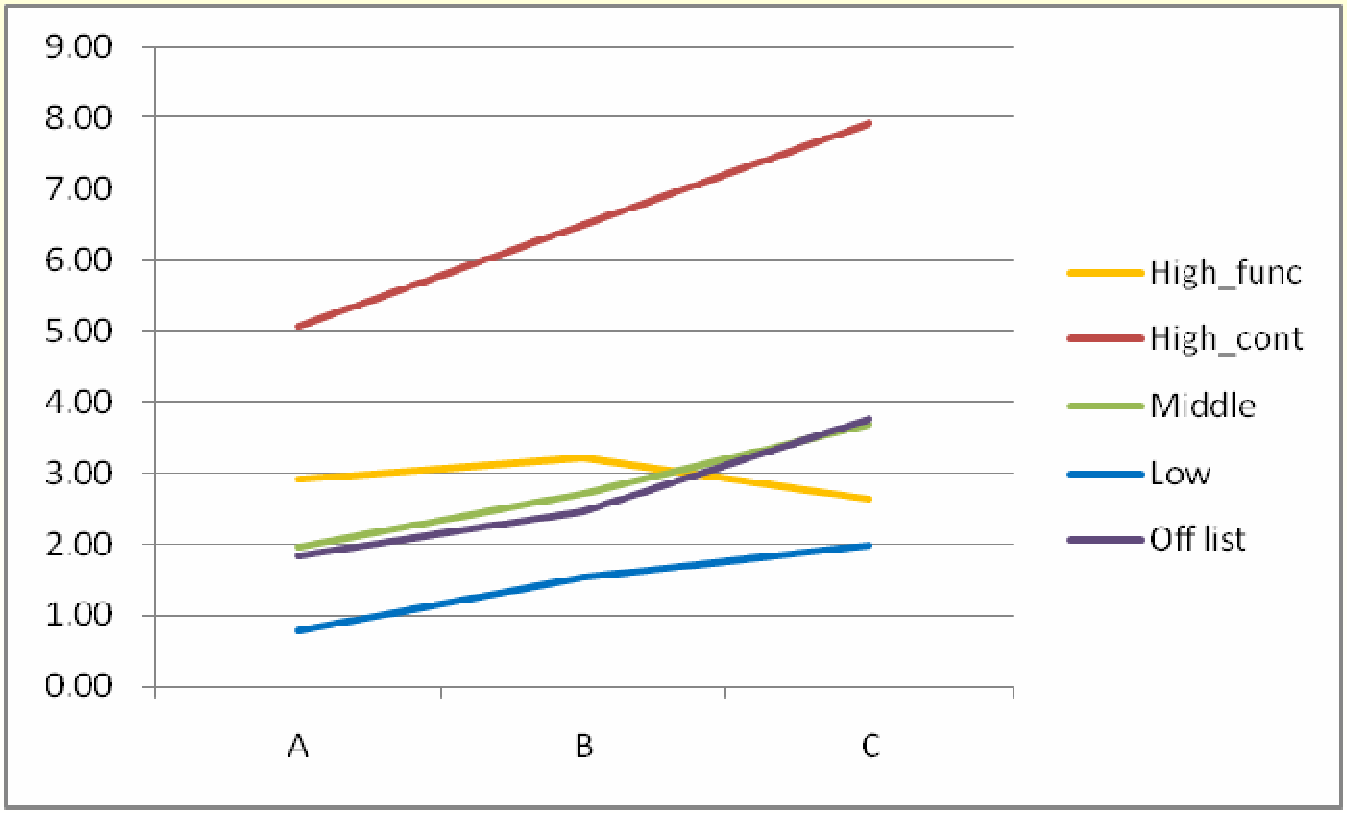


- A < B < C
- Remarkable increase (High) B → C
- Few low freq word tokens

4. Result

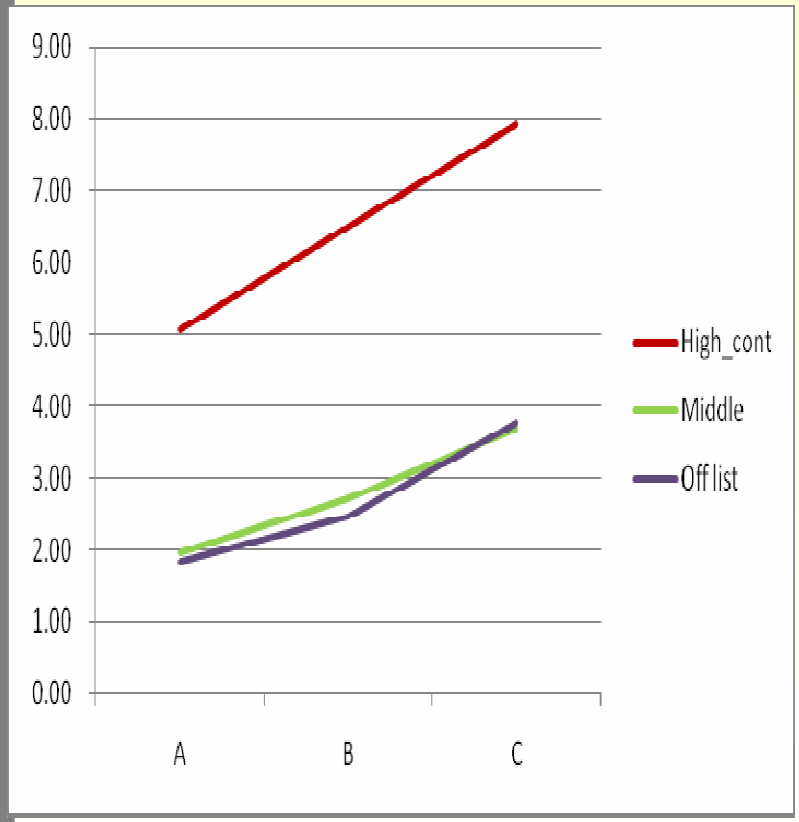
Comparison among groups

- Average of diversity of each group in frequency layers



4. Result

Comparison among groups



significant among all groups

- High freq (content words)

  A < B <C

  [$F(2,25)=21.484$, $p<.001$]

- Middle freq

  A < B < C

  [$F(2,25)=23.501$, $p<.001$]

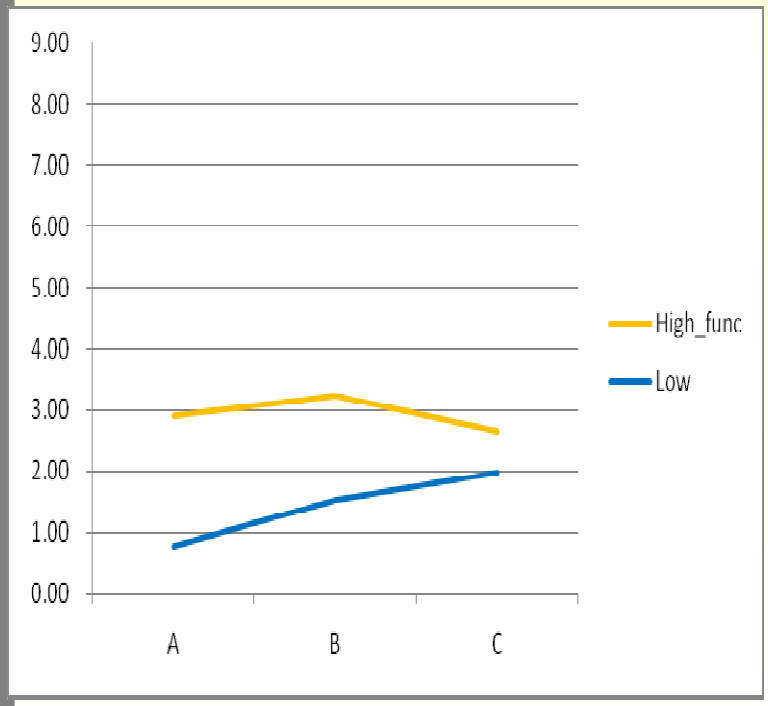
- Off list freq

  A < B < C

  [$F(2,25)=27.020$, $p<.001$]

4. Result

Comparison among groups



Not significant among all groups

- High freq (function words)

  B > C

[*F*(2,25)=4.375, *p*<.05]

- Low freq

  A < B, A< C

[*F*(2,25)=8.258, *p*<.01]

4. Result
Correlation

	High_F	High_C	Middle	Low	Off list
High_F	1				
High_C	−0.059	1			
Middle	−0.210	**0.706**	1		
Low	0.125	0.279	0.427	1	
Off list	−0.167	0.543	0.670	0.453	1

- **Strong correlation**
  → High (content words) ⇔ Middle
- **Moderate correlation**
  →High (content words) ⇔ Off list
  →middle ⇔Low
  → middle ⇔Off list
  → low ⇔Off list

4. Result
Correlation

	High_F	High_C	Middle	Low	Off list
High_F	1				
High_C	−0.059	1			
Middle	−0.210	0.706	1		
Low	0.125	0.279	0.427	1	
Off list	−0.167	0.543	0.670	0.453	1

□ No correlation

High (function) ⇔ others

4. Summery of result

- High (function words)
- High (content words)
- Middle
- Low
- Off list

} × Diversity

4. Summery of result

- High (function words)
- **High (content words)**
- **Middle**
- Low
- **Off list**

× Diversity

Diversity of high frequency content words
middle frequency words
Off list words

Interrelation

→ **categorize Japanese learners**

5. Discussion and conclusion

- Too few participants, Small corpus
- Qualitative analysis
- Comparison the result with judgment by evaluators
- Comparison b/w NS and learners
- Learnability of lower freq. words

References

- Bulté, B., A. Housen, M. Pierrard. and S. van Daele (2008)"Investigating Lexical Proficiency Development over Time- the case of Dutch-speaking Learners of French in Brussels." *French Language Studies.* Vol.18. pp. 277-298.
- Daller, J., J. Milton and J. Treffers-Daller (eds.) (2007) *Modeling and Assessing Vocabulary Knowledge*. Cambridge University Press.
- David, A. (2008) "A developmental Perspective on Productive Lexical Knowledge in L2 Oral Interlanguage". *French Language Studies.* Vol.18. pp. 315-331.
- Laufer, B. and Nation, P. (1995) "Vocabulary Size and Use: Lexical Richness in L2 Written Production". *Applied Linguistics 16/3*. pp.307-322.
- Malvern, D. and Richards, B. (2009) "A New Method of Measuring Rare Word Diversity: the Example of L2 Learners of French." In Richard, B., H. Daller., D. Malvern., P. Meara, J. Milton and J. Treffers-Daller. (eds) *Vocabulary Studies in First and Second Language Acquisition- the Interface between Theory and Application-*. Palgrave Macmillan. pp.164-178.
- Milton, J. (2009) *Measuring Second Language Vocabulary Acquisition*. Multiligual Matters.
- 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*. pp.107-125

References

- Read, J. (2000) *Assessing Vocabulary*. Cambridge University Press.
- Read, J. (2005) Applying Lexical Statistics to the IELTS Speaking Test. *Research Notes*. vol.20. pp.12-15
- Richard, B., H. Daller., D. Malvern., P. Meara, J. Milton and J. Treffers-Daller. (eds) (2009) *Vocabulary Studies in First and Second Language Acquisition- the Interface between Theory and Application-*. Palgrave Macmillan.
- Thomas, A. (2008) La Mesure des Progrès Lexicaux en FL2 Avancé. In J. Durand., B. Habert., and B. Laks. (eds) *Congrès Mondial de Linguistique Français- CMLF08*. Institut de Linguistique Français. pp. 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?". In Daller, J., J. Milton and J. Treffers-Daller (eds.) *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press. pp.133-149.
- Treffers-Daller, J. (2009) "Language Dominance and Lexical Diversity : How Bilinguals and L2 learners Differ in their Knowledge and Use of French Lexicaland Functional Items". In Richard, B., H. Daller., D. Malvern., P. Meara, J. Milton and J. Treffers-Daller. (eds) *Vocabulary Studies in First and Second Language Acquisition- the Interface between Theory and Application-*. Palgrave Macmillan. pp. 74-90.

References

- Verlinde, S. and Selva, T. (2001) "Corpus-Based versus Intuition-Based Lexicography: Defining a Word List for a French Learner's Dictionary". In P. Rayson, A., T, Wilson., A. McEnery. and S. Khoja. (dir.) *Proceedings of the Corpus Linguistics 2001 conference*. Technical Papers 13. pp.594-598.
- Van Hout, R. and Vermeer, A. (2007) "Comparing Measures of Lexical Richness". In Daller, J., J. Milton and J. Treffers-Daller (eds.) *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press. pp.93-114

Acknowledgement

- This research was supported by a grant for the Global COE Program, "Corpus-based Linguistics and Language Education", from the Ministry of Education, Culture, Sports, Science and Technology of Japan.