

COMPARAISON DE L'USAGE DES "MULTI WORD UNITS" ENTRE LES APPRENANTS JAPONAIS DU FRANÇAIS ET LES FRANCOPHONES NATIFS

KAORI SUGIYAMA

(TUFS&U. DE BORDEAUX 3)

IPFC2012-PARIS
10/12/2012

PLAN

- 1. Objectifs de l'exposé**
- 2. Qu'est-ce que les MWUs?**
- 3. Questions de la recherche**
- 4. Méthode de recherche**
- 5. Résultats et analyse**
- 6. Problèmes et perspectives**

PLAN

- 1. Objectifs de l'exposé**
2. Qu'est-ce que les MWUs?
3. Questions de la recherche
4. Méthode de recherche
5. Résultats et analyse
6. Problèmes et perspectives

1. OBJECTIFS DE L'EXPOSÉ

- **description linguistique de l'usage des «multi-word units»**
 - Français natifs vs Japonais
 - Débutants vs Avancés

} **Contrastive Interlanguage Analysis (Granger 1996)**

- **analyse globale et objective**
 - Analyse basée sur le corpus
 - Statistiques → tester l'intuition, reconsidérer l'intuition

- **utilisation des corpus IPFC en dehors des études phonético-phonologiques**

PLAN

1. Objectifs de l'exposé
- 2. Qu'est-ce que les MWUs?**
3. Questions de la recherche
4. Méthode de recherche
5. Résultats et analyse
6. Problèmes et perspectives

2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

amalgams	gambits	preassembled speech prefabricated routines and patterns
automatic chunks	gestalt holistic	ready-made expressions ready-made utterances
clichés	holophrases	recurring utterances
co-ordinate constructions	idiomatic	rote
collocations	idioms	routine formulae
composites	irregular	schemata
conventionalized forms	lexical(ized) phrases	semi-preconstructed phrases that constitute single choices
F[ixed] E[xpressions] including I[dioms]	lexicalized sentence stems	sentence builders
fixed expressions	multiword units	stable and familiar expressions with specialized subsenses
formulaic language	non-compositional	stereotyped phrases
formulaic speech	non-computational	stereotypes
formulas/formulae	non-productive	stock utterances
fossilized forms	non-propositional	synthetic unanalysed chunks of speech
frozen metaphors	petrifications	
frozen phrases	praxons	

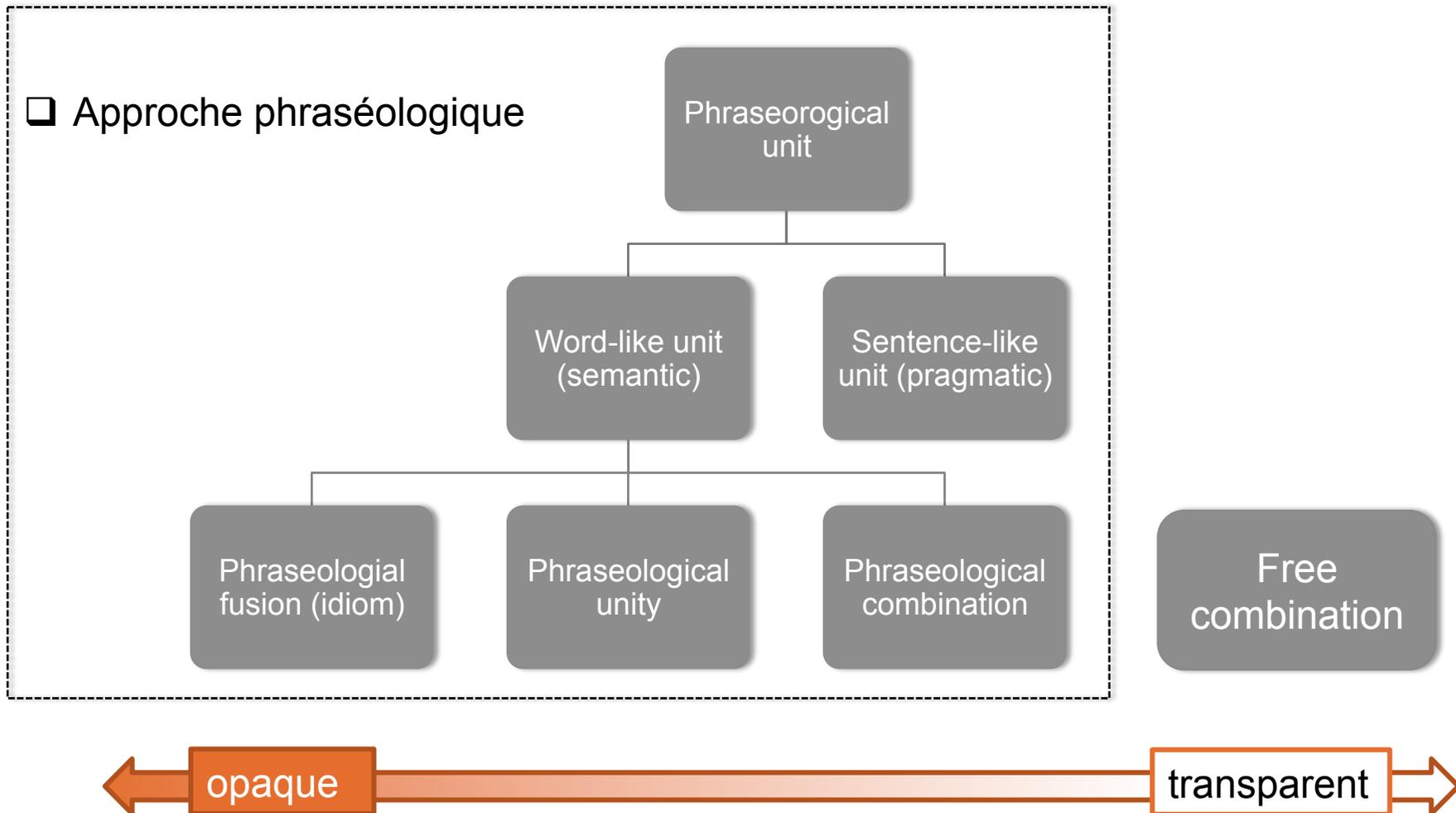
Dénominations des
mots consécutifs

Wray (2000)

2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

- ❑ **Inconformité des dénominations des « mots consécutifs »**
 - ⇒ **opposition de 2 approches (Nesselhauf 2004)**
 - Approche phraséologique
 - Approche basée sur la fréquence

2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?



Cowie (1998), Vinogradov (1947, cité dans Cowie (ibid.))

2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

□ Approche basée sur la fréquence

- 1) Formes des mots (× lemmes)
- 2) Séquences des formes (× aspects grammatico-syntaxiques)
- 3) Séquences des formes fréquentes

(Biber 2009)

□ Types d'analyses

- 1) Collocation
- 2) N-grammes

(Granger and Paquot 2008)

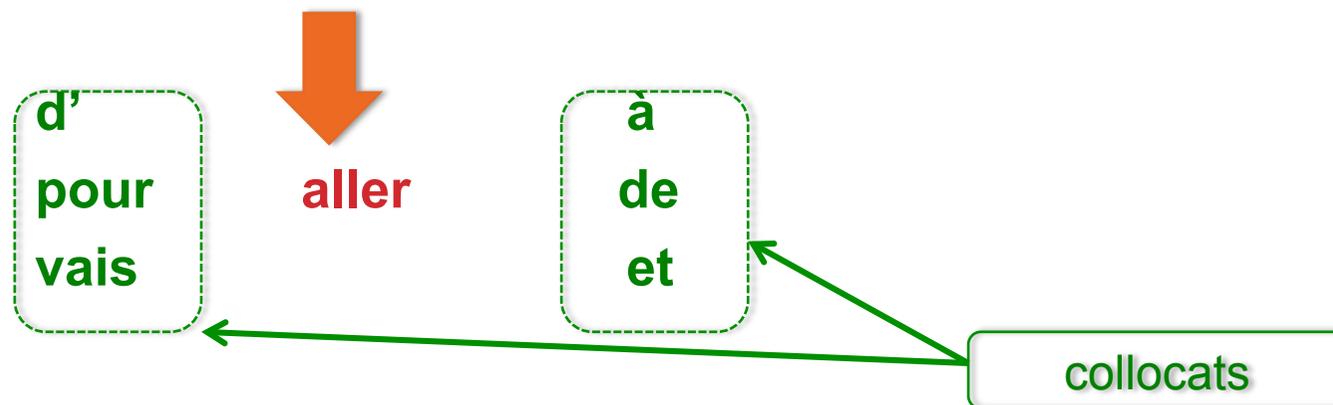
2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

□ Collocation (Biber et al. 1989)

- analyses de séquences statistiquement fréquentes
- Unité = nœud (node word) + collocat

→ sélectionné par le **jugement linguistique préalable**

ex. « aller » → nœud



2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

- ❑ N-grammes (Granger and Paquot 2008)
 - extraits à l'aide d'un traitement automatique
 - séquences de mots

Ex:

Il aime la musique.

2 grammes

2. QU'EST-CE QUE LES « MULTI-WORD UNITS »?

- ❑ N-grammes (Granger and Paquot 2008)
 - extraits à l'aide d'un traitement automatique
 - séquences de mots

Ex:

Il aime la musique.

3 grammes

PLAN

1. Objectifs de l'exposé
2. Qu'est-ce que les MWUs?
- 3. Questions de la recherche**
4. Méthode de recherche
5. Résultats et analyse
6. Problèmes et perspectives

3. QUESTIONS DE LA RECHERCHE

□ N-grammes « sous-utilisés »

⇒ manque de « native-likeness »

- Caractéristiques des N-grammes « sous-utilisés »?
- Relation entre la différence des N-gramme et la différence de niveaux de compétence linguistique?

PLAN

1. Objectifs de l'exposé
2. Qu'est-ce que les MWUs?
3. Questions de la recherche
- 4. Méthode de recherche**
5. Résultats et analyse
6. Problèmes et perspectives

4. MÉTHODE DE RECHERCHE

- ❑ **CIA (Granger 1996) :**
Comparaison l'IPFC-JP avec le corpus natif
⇒ corpus multilingue du français (Aix)

- ❑ **IPFC-JP**
 - Répétition
 - Prononciation à partir d'une liste de mots
 - Lecture à haute voix
 - Conversation guidée
 - Conversation libre

- ❑ **Participants de l'IPFC-JP**
 - Débutants (2^{ème} année de licence)~
Avancés (doctorants/séjour dans un pays francophone)
 - 38 apprenants du français (conversation libre)

4. MÉTHODE DE RECHERCHE

□ Taille des corpus

- corpus de l'étude = IPFC-JP: 34533 mots
- corpus de référence = Corpus Natif: 65591 mots

□ Transcription

- Hésitation (bah, beh, ben) ⇒ HESIT
- Back-channel (mmm, hum) ⇒ BC
- Nom propre ⇒ NP
- Interjection (ah, oh) ⇒ INT
- Interruption ⇒ supprimé
- Mot étranger ⇒ FW (apprenants seuls)

4. MÉTHODE DE RECHERCHE

- ❑ Sous-utilisations de N-grammes (N=2,...,5)
- ❑ Extraction des sous-utilisations de N-grammes
WordSmith ver.6
- ❑ « sous-utilisation »= **Keyness** (log-vraisemblance)
sous-utilisation ≤ -20

4. MÉTHODE DE RECHERCHE

❑ Liste de keyness

comparaison : natifs ↔ apprenants

avancés ↔ débutants

❑ Niveaux de compétence linguistique

➤ **Avancé (13 pers., 21534 mots)**

⇒ masters, doctorants

ayant vécu dans un pays francophone (min. 1an)

➤ **Débutant-intermédiaire (25 pers., 12999mots)**

⇒ licences, masters, doctorants

pas d'expérience de longue durée

PLAN

1. Objectifs de l'exposé
2. Qu'est-ce que les MWUs?
3. Questions de la recherche
4. Méthode de recherche
- 5. Résultats et analyse**
6. Problèmes et perspectives

5. RÉSULTATS ET ANALYSES

Key word	freq. in JP	freq. in FR	Keyness
TU VOIS	7	128	-68,19
TU SAIS	6	120	-66,11
QU ON	6	111	-59,39
NON MAIS	2	71	-46,01
OUAIS BC	5	85	-43,96
ILS ONT	7	94	-43,58
QUE ÇA	4	78	-42,56
MAIS BON	2	65	-41,28
QUAND MÊME	12	102	-35,15
VRAI QUE	3	62	-34,54
QUI SONT	2	56	-34,25
EN A	2	56	-34,25
CE QUI	3	61	-33,79
EN FAIT	50	217	-32,58
TOUT ÇA	2	50	-29,61
QU ILS	5	60	-26,16
QUI EST	14	90	-23,79
DANS LES	2	42	-23,52
DE FAIRE	1	36	-23,39
ENFIN JE	1	36	-23,39
QU IL	28	132	-22,92
JE SAIS	50	188	-20,84
EST QUE	2	38	-20,53
ET TOUT	4	47	-20,24

Key word	freq. in JP	freq. in FR	Keyness
EST VRAI QUE	3	61	-33,79
Y EN A	2	46	-26,55
IL Y EN A	2	45	-25,79
C EST VRAI QUE	3	48	-24,18

5. RÉSULTATS ET ANALYSES

Key word	freq. in JP	freq. in FR	Keyness
TU VOIS	7	128	-68,19
TU SAIS	6	120	-66,11
QU ON	6	111	-59,39
NON MAIS	2	71	-46,01
OUAIS BC	5	85	-43,96
ILS ONT	7	94	-43,58
QUE ÇA	4	78	-42,56
MAIS BON	2	65	-41,28
QUAND MÊME	12	102	-35,15
VRAI QUE	3	62	-34,54
QUI SONT	2	56	-34,25
EN A	2	56	-34,25
CE QUI	3	61	-33,79
EN FAIT	50	217	-32,58
TOUT ÇA	2	50	-29,61
QU ILS	5	60	-26,16
QUI EST	14	90	-23,79
DANS LES	2	42	-23,52
DE FAIRE	1	36	-23,39
ENFIN JE	1	36	-23,39
QU IL	28	132	-22,92
JE SAIS	50	188	-20,84
EST QUE	2	38	-20,53
ET TOUT	4	47	-20,24

Key word	freq. in JP	freq. in FR	Keyness
EST VRAI QUE	3	61	-33,79
Y EN A	2	46	-26,55
IL Y EN A	2	45	-25,79
C EST VRAI QUE	3	48	-24,18

Lexiques
« en »

5. RÉSULTATS ET ANALYSES

Key word	freq. in JP	freq. in FR	Keyness
TU VOIS	7	128	-68,19
TU SAIS	6	120	-66,11
QU ON	6	111	-59,39
NON MAIS	2	71	-46,01
OUAIS BC	5	85	-43,96
ILS ONT	7	94	-43,58
QUE ÇA	4	78	-42,56
MAIS BON	2	65	-41,28
QUAND MÊME	12	102	-35,15
VRAI QUE	3	62	-34,54
QUI SONT	2	56	-34,25
EN A	2	56	-34,25
CE QUI	3	61	-33,79
EN FAIT	50	217	-32,58
TOUT ÇA	2	50	-29,61
QU ILS	5	60	-26,16
QUI EST	14	90	-23,79
DANS LES	2	42	-23,52
DE FAIRE	1	36	-23,39
ENFIN JE	1	36	-23,39
QU IL	28	132	-22,92
JE SAIS	50	188	-20,84
EST QUE	2	38	-20,53
ET TOUT	4	47	-20,24

Key word	freq. in JP	freq. in FR	Keyness
EST VRAI QUE	3	61	-33,79
Y EN A	2	46	-26,55
IL Y EN A	2	45	-25,79
C EST VRAI QUE	3	48	-24,18

Syntaxe

« que »

- Conjonction de subord.
- Locution conj. de subord.
- Pronom relatif
- Interrogatif
- Mise en relief

« qui »

- Pronom relatif
- Interrogatifs
- Mise en relief

5. RÉSULTATS ET ANALYSES

Key word	freq. in JP	freq. in FR	Keyness
TU VOIS	7	128	-68,19
TU SAIS	6	120	-66,11
QU ON	6	111	-59,39
NON MAIS	2	71	-46,01
OUAIS BC	5	85	-43,96
ILS ONT	7	94	-43,58
QUE ÇA	4	78	-42,56
MAIS BON	2	65	-41,28
QUAND MÊME	12	102	-35,15
VRAI QUE	3	62	-34,54
QUI SONT	2	56	-34,25
EN A	2	56	-34,25
CE QUI	3	61	-33,79
EN FAIT	50	217	-32,58
TOUT ÇA	2	50	-29,61
QU ILS	5	60	-26,16
QUI EST	14	90	-23,79
DANS LES	2	42	-23,52
DE FAIRE	1	36	-23,39
ENFIN JE	1	36	-23,39
QU IL	28	132	-22,92
JE SAIS	50	188	-20,84
EST QUE	2	38	-20,53
ET TOUT	4	47	-20,24

Key word	freq. in JP	freq. in FR	Keyness
EST VRAI QUE	3	61	-33,79
Y EN A	2	46	-26,55
IL Y EN A	2	45	-25,79
C EST VRAI QUE	3	48	-24,18

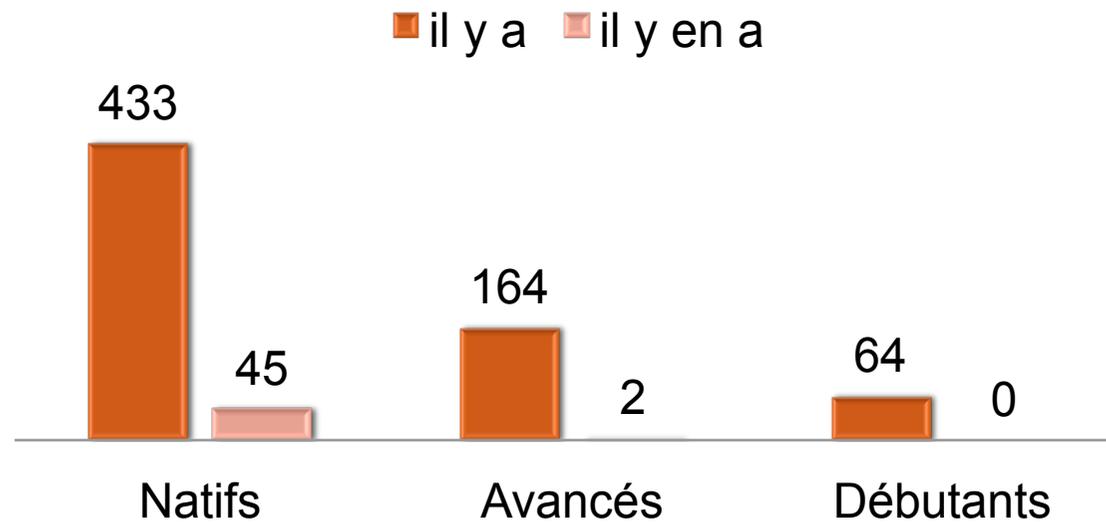
Discursif

- Indication de structure de l'interaction (ex: mais bon)
- Manifestation de co-construction (ex: tu vois, tu sais, quand même)
- Marqueur de production discursive (ex: en fait, enfin je)

(Bruxelles et al. 2001)

5. RÉSULTATS ET ANALYSES

☐ Lexique « en »



5. RÉSULTATS ET ANALYSES

Natifs

il y en a un qui a dû ouvrir (N12)

et #HESIT# donc il y en a qui sont #HESIT# qui sont vachement bien (N18)

Apprenants

➤ Avancés

il y en a il y en avait rien là bas (A9)

trois semestre oui c est ça dans la troisième semestre il y en a quinze (A7)

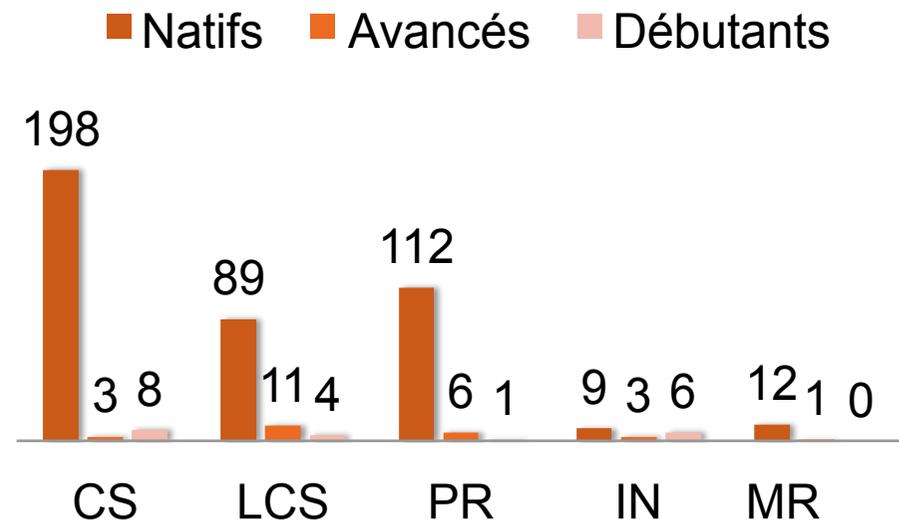
➤ Débutants

(pas de données)

5. RÉSULTATS ET ANALYSES

☐ Syntaxe : « que »

		Natifs	Avancés	débutants
qu'on	CS	56	0	0
	LCS	26	3	0
	PR	43	3	0
	IN	0	0	0
	MR	4	0	0
que ça	CS	37	1	0
	LCS	19	1	0
	PR	8	0	0
	IN	0	1	0
	MR	0	1	0
qu'ils	CS	34	0	2
	LCS	9	2	0
	PR	27	1	0
	IN	3	0	0
	MR	4	0	0
qu'il	CS	71	2	6
	LCS	35	5	4
	PR	34	2	1
	IN	6	2	6
	MR	4	0	0



- CS= conj. de subord.
- LCS= locution conj. de subord.
- PR= pronom relatif
- IN= interrogatif
- MR= mise en relief

5. RÉSULTATS ET ANALYSES

☐ Natifs

avoir des ouvertures sur tous ce qui est métiers de la communication donc il faudrait qu'il y ait des options (N3)

parce qu'on avait monté là bas le un #NP# (N11)

le tout c est de savoir ce qu'on va en faire ouais (N14)

c est du céleri qu'ils mangent là bas (N18)

☐ Apprenants

➤ Avancés

voilà donc je pense que ça vient de là (A2)

c est ce qu'on a acheté parce qu'on a déjà plein de trucs (A7)

➤ Débutants

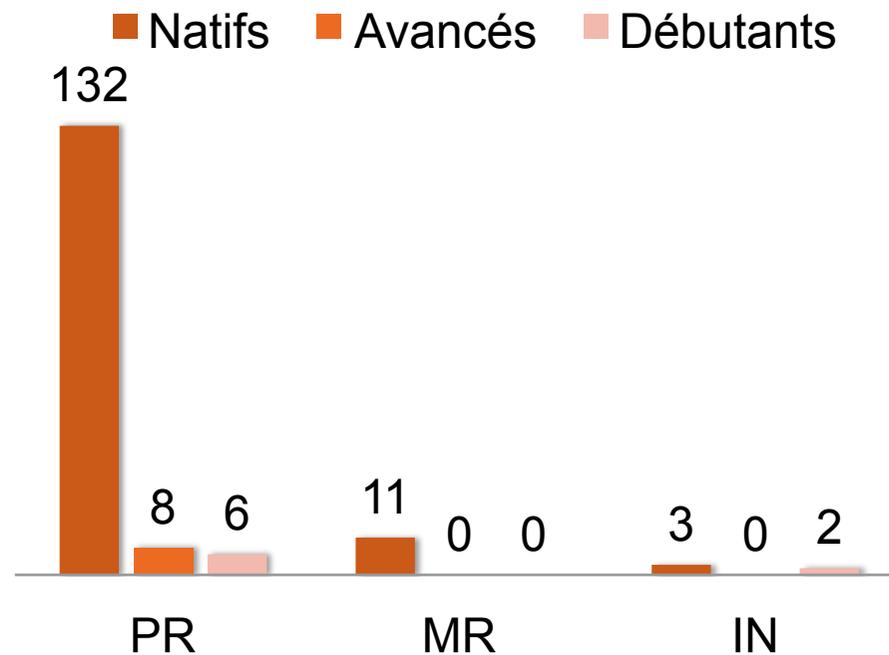
parce qu'il y a des joueurs préférés (D 24)

mon oncle dit dit que #HESIT# #HESIT# #HESIT# dit #HESIT# qu'il a été volé qu'il a été volé #HESIT# (D25)

5. RÉSULTATS ET ANALYSES

☐ Syntaxe « qui »

		Natifs	Avancés	débutants
qui sont	PR	49	2	0
	MR	7	0	0
	IN	0	0	0
qui est	PR	83	6	6
	MR	4	0	0
	IN	3	0	2



5. RÉSULTATS ET ANALYSES

☐ Natifs

et ce qui est intéressant c est qu il y a des étudiants (N14)

ce sont des conventions qui sont passées (N8)

☐ Apprenants

➤ Avancés

mais en fait le vin suisse est très très bon mais ce qui est le problème est (A8)

j ai j ai un ami à à #NP# #HESIT#
qui est japonais (A4)

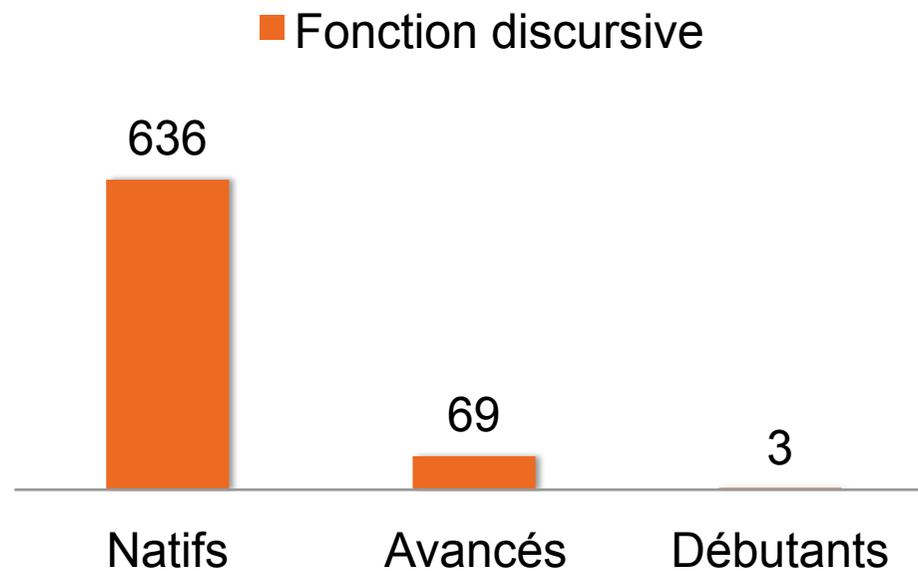
➤ Débutants

qui est #NP# est qui est qui qui c est qui (D13)

5. RÉSULTATS ET ANALYSES

☐ Discursif

	Natifs	Avancés	débutants
mais bon	65	2	0
tu vois	117	5	0
tu sais	98	2	1
quand même	102	11	0
en fait	218	48	2
enfin je	36	1	0



5. RÉSULTATS ET ANALYSES

☐ Natifs

tu pars avec tes parents tu fais plein de voyages à quinze ans tu vois j en ai marre (N3)

mais #HESIT# j ai un avantage aussi #HESIT# en même temps c est un peu un inconvénient enfin je sais pas (N5)

Je t assure Parfois j ose pas aller à la salle de bain tu sais quand il fait nuit #INT# (N17)

☐ Apprenants

➤ Avancés

c est c est un boulangerie de bio et un peu de luxe que que les japonais adorent tu vois et donc on est on est allé à dans ce ce boulangerie là (A7)

c est #INT# j ai raté un un seul cours quand même mais (A10)

➤ Débutants

c était trop trop trop difficile en fait j ai pas fini donc (D19)

PLAN

1. Objectifs de l'exposé
2. Qu'est-ce que les MWUs?
3. Questions de la recherche
4. Méthode de recherche
5. Résultats et analyse
- 6. Problèmes et perspectives**

6. PROBLÈMES ET PERSPECTIVES

- ❑ Étudiants de TUFS ≠ Représentatif des apprenants Japonais??
- ❑ Taille de corpus
- ❑ Approfondissement des recherches sur chacune des composantes de la sous-utilisation

BIBLIOGRAPHIE

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. (1989) *Longman Grammar of Spoken and Written English*. Longman.

Biber, D. (2009) A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* (14). pp. 275-311.

Bruxelles, S. et Traverso, V. (2001) Ben: apport de la description d'un « petit mot » du discours à l'étude des polylogues. *Marges Linguistiques* 2. pp. 38-55.

Cowie, A. (ed.) (1998). *Phraseology: Theory, analysis, and applications*. Oxford University Press.

Granger, S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora, In. Aijmer K., Altenberg B. and Johansson M. (eds.) *Languages in Contrast. Text-based cross-linguistic studies*. Lund University Press. pp. 37-51.

Granger, S. and Paquot, M. (2008) Disentangling the phraseological web. In. Granger, S. and Meunier, F. (eds.), *Phraseology: An interdisciplinary perspective*. John Benjamins. pp.27-49.

Nesselhauf, N. (2004) *Collocation in a learner corpus*. Benjamins.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(3), pp.463-489.

MERCI POUR VOTRE ATTENTION

CETTE RECHERCHE EST SOUTENUE PAR
L'**ITP-EUROPE** DE L'UNIVERSITÉ DES
LANGUES ETRANGÈRES DE TOKYO.