Relative frequencies of academic phraseology for writing in the ELF scope

Alejandro Curado Fuentes

Cáceres, University of Extremadura

December 2012

Zaragoza conference

English as a Scientific and Research Language

Relative frequencies of academic phraseology for writing in the ELF scope

Three main parts:

1. Academic writing setting
2. The academic writing corpus analysis
3. The evaluation of phraseology for writing

The setting

Writing ESP :
EST / EAP / EFL
(tertiary levels)

From "words and structures" > "texts and purposes" > "learners and genres" > "contexts and interactions" (Upton 2012)

Our context? = writing / speaking?

The setting

Writing in EAP / EST = ELF?

"Well-qualified academics" (faculty & researchers / graduate students) writing in ELF settings (journals, project reports,… blogs, e-mails, social networks…)

(Mauranen, 2012)

The setting

Lingua Franca as opportunity?

--EFL countries (Brazil, France, Spain...) >

(ESP as 'butler' of Literature department 'dinosaurs'–
Raimes, 1991; Aguado & Curado, 2012; Hyland, 2012)

--Academic literacies as naturally demanded in practice, i.e., students' own perceptions of genre / text / phrases...
+ intercultural exchanges / acculturation (Pérez Llantada, 2012)

--Writing in the Humanities versus Empirical Sciences? (Mauranen, 2012)

The setting

Academic / scientific phraseology for writing: our case scenario (B1 / B2 levels)

Very different frequencies, notions, and meanings across disciplines (cf. Hyland, B. 2009; Durrant, 2009; e.g.;

1) 80% Engineering; education, Veterinary science, then, Business, Law, Philologies, History

English linguistics (+ textual ref.) Softer / Humanities (internationalization?)

Electrical engineering (research oriented.)

(Hard sciences vs.)

(Medicine > in Badajoz but not many)

The corpus analysis:
Relative wordlist and co-occurence frequencies
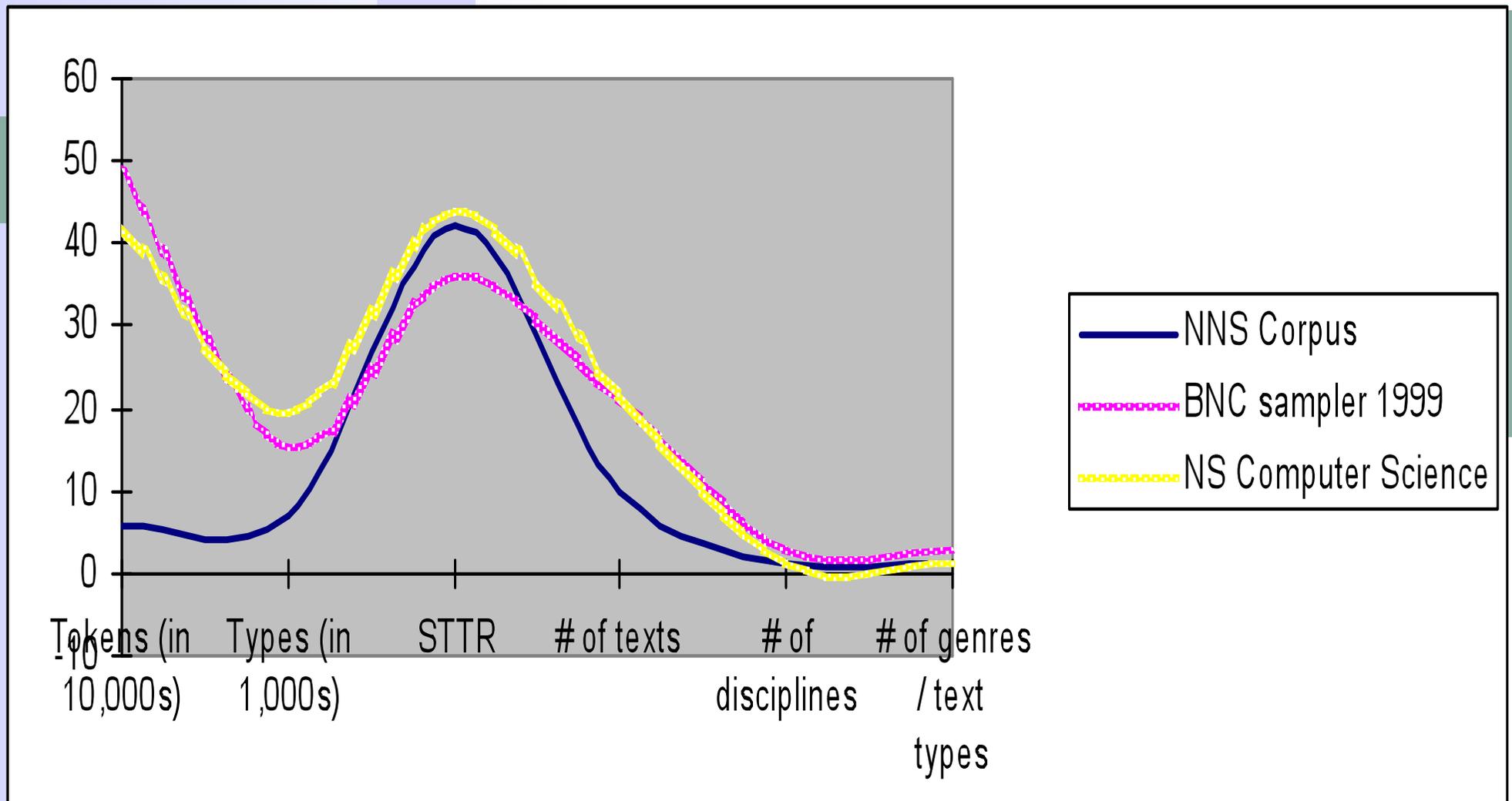
Relative wordlist frequencies (% & positions w/i corpus)

WE WORD	NNS Corpus for Case Study	BNC selection	C.S. NS (40,180 tokens)
	>> (+15)		Is
			For
			That
HAVE	>		Be
IN	> (+5)		Are
CAN	>>		As
			With
TO		<	This
AT	>		By
			On
FOR	>		It >
USE	>>		From
AS	>		Was >
FROM		<	Can
THAT	>		Not >
WHICH	=		Which
IS		<	Have
BUT	>		We <<
ON	>		Within
			These
			At
			Were
			Also

The corpus analysis
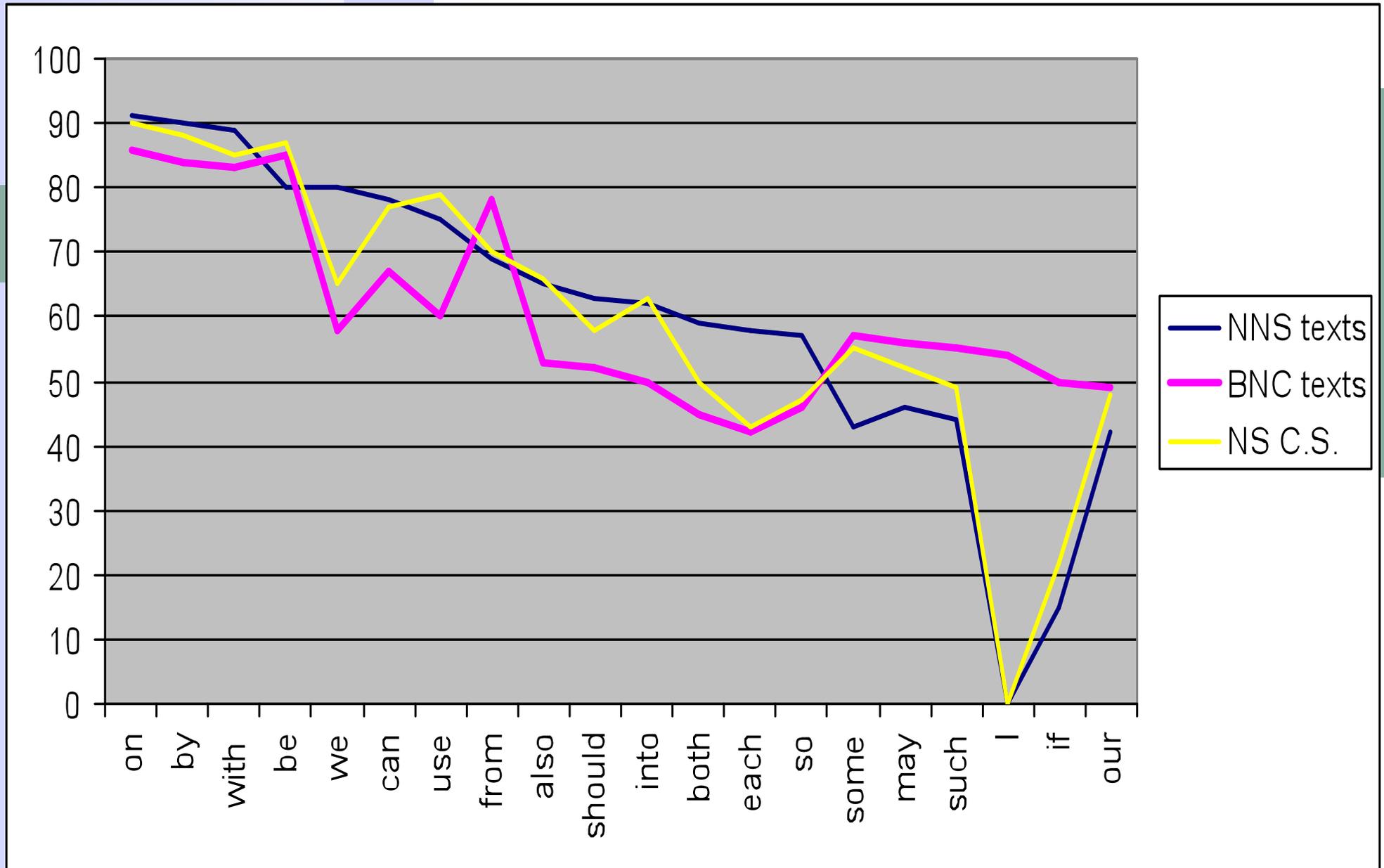
PHRASEOLOGICAL UNITS	Similar Use (less than 15% difference)	NS (more than 15% for NS)	NNS (more than 15% for NNS)
Words co-occurring together across texts (collocations)	Appear* + to be (20 / 20.4 / 19%)	It is possible to (28 / 8 / 28.4%)	We observe that (0 / 14.7 / 0%)
Grammatical aspect involved in collocation (colligations)	Form (v) + basis for + (NP) (26.3 / 17.6 / 21%)	Noun + *to* + NP (e.g., *key to, access to, way to …* (26.5 / 1.2 / 17%)	Be + asked to (present tense) (0 / 61.5 / 13%)
A semantic set associated to word (Semantic Associations)	In the field of + [area] (20 / 11.5 / 16%)	SEEK + [legal requirement] (28 / 0 / 18%)	Related to + [functionability] (6 / 76.9 / 36%)
3-6 word bundle in specific positions in text (textual colligates)	As a result of (beg. paragraphs) (20 / 31.5 / 26%)	One of the most + adj. (beg. sentences) (23.2 / 4.3 / 9%)	For this reason, (beg. sentences) (22.9 / 43 / 13%)

The evaluation

Questions for ELF writing

>NNS relative frequencies can be valid for publication?

>Native-like versus ELF?

Frequencies according to external factors

Collocation	*Such as* + examples (52 / 56% --C.S. papers) [research writing]	*If and only if* (71.4% --BNC: Logic) [topic / subject]
Colligation	*I had* + past participle (47% --BNC reports) [genre]	*is* + *to be* + past participle (22 / 17.8% -- networking procedures) [text type]
Semantic Association	*Be* + *applied to* + operation (17 / 25.6 % --C.S. papers > Methodology)	*Be / appear* + *on the right* + *side* (19 / 26.6% --C.S.: papers)
Textual Colligation	*There is no* + noun (beg. paragraphs) (34.8% -- BNC articles)	*This form* + *be completed* (beg. paragraphs) (16.4% -- BNC: surveys)

Register as reference

Example: We + observe + that (NNS)

NP + may be observed (COCA)

ACAD: Modality & passive (Biber, 2006)

Maybe little explored in NS vs. NNS for academic stance: (Peacock, 2010)

	WORD/PHRASE	TOKENS ACADEMIC	TOKENS NEWS	PM 1	PM 2	RATIO
1	CAN BE GENERALIZED	109	0	1.20	0.00	119.69
2	CAN BE ESTIMATED	112	1	1.23	0.01	112.80
3	MUST BE INTERPRETED	99	1	1.09	0.01	99.70
4	CAN BE ILLUSTRATED	95	1	1.04	0.01	95.68
5	MAY BE UNDERSTOOD	85	0	0.93	0.00	93.34

Modals with passive in news

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	MAY BE EDITED	119	0	1.30	0.00	129.75
2	WILL BE UNVEILED	38	0	0.41	0.00	41.43
3	'D BE LYING	34	1	0.37	0.01	33.76
4	WILL BE TELEVISED	34	1	0.37	0.01	33.76
5	WILL BE SHUT	32	1	0.35	0.01	31.77
6	'LL BE PLAYING	29	0	0.32	0.00	31.62
7	WILL BE DECORATED	26	0	0.28	0.00	28.35
8	WILL BE PRICED	28	1	0.31	0.01	27.80
9	COULD BE HURT	25	0	0.27	0.00	27.26

BE U

Academic Register relative frequencies (NS)

Results / findings + can be generalized + to + other + NN
(109 times = 0.05% for CAN)

We can generalize + NN (results, text…) =
(9 times = 0.004)

NN + *could be predicted + from…*
(44 times = 0.04% for COULD)

We could predict (4 times = 0.004)

Other frequencies
(BAWE = Academic L1 vs. L2 ?)

We could
Hits: 267 (32.0 per million) = 2.8% for COULD

Could + BE + p.p. (PASSIVE)
Hits: 4,452 (534.1 per million) = 37.8% for COULD

It could + be + argued + that
(87.2% = English / 2.2% = Slovenian--level 3 /
critique) / Hindi / 0.3% …)
[81 times = 2.3% for ARGUE]

we could argue that
(100% English > concluding / cause-effect)
[2 times = 0.05% of ARGUE]

Academic L1 vs L2?

We could + find
42% = Mandarin Chinese > essays & business report /
42% = English > recount method, discussions >
[11 times = 0.15% for FIND vs. 0.33]

We could + see
45% = Japanese > essays & exercises /
5.5% = English = subordinate
it is possible that we could see…)
[10 times = 0.01% for SEE vs. 0.4]

Comparing with other academic corpora (e.g., Hong Kong Polytechnic RAs)
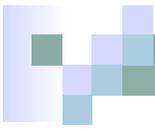
we could + argue = **None**
we could + see = **None**

It could + be + argued + that (T-score = 3.14)

We could + find = 2.4 (never followed by THAT)
(15.3% in a negative clause)

NP + *could + BE + seen + as a* + NP = T-score is 4

****NP + *could + BE + found + in* + PLACE / TIME PERIOD/ AREA** = 7 times but t-score = 0.4

**Caution about overuse of "anything"
(Kobayashi & Tanaka, 2010)

Evaluating academic phraseology production
(60 students: 30 B1 / 30 B2)

Eng (ther one write mat conte	EXAMPLES OF USE IN SENTENCES OR EXPRESSIONS (if more than one meaning in the word, please give more than one example)	WORD (Across disciplines —COCA academic families)	TRANSLATION	text(s) / ere you the (journal stract, eport, tion, , ...?)
1.		DEVELOPMENT		
2.		RESEARCH		
3.		PROVIDE		
4.		INCREASE		
		LEVEL		

Evaluation

1) "Academic" errors: Lack of suitability for context / text (e.g., *reliability can't be proved* in article)

2) "L1" errors: Thinking in Spanish (e.g., *he affirmed that the problem could be solved by another different method*)

3) Level errors: Grammar & lexis (e.g., *it was detected failure on that*)

4) Spanish translation errors (e.g., *discutimos el asunto durante una hora*)

5) "Good" choices: using relative frequencies in context (e.g., *In other words, I would like to emphasize that …* in presentations/discussions)
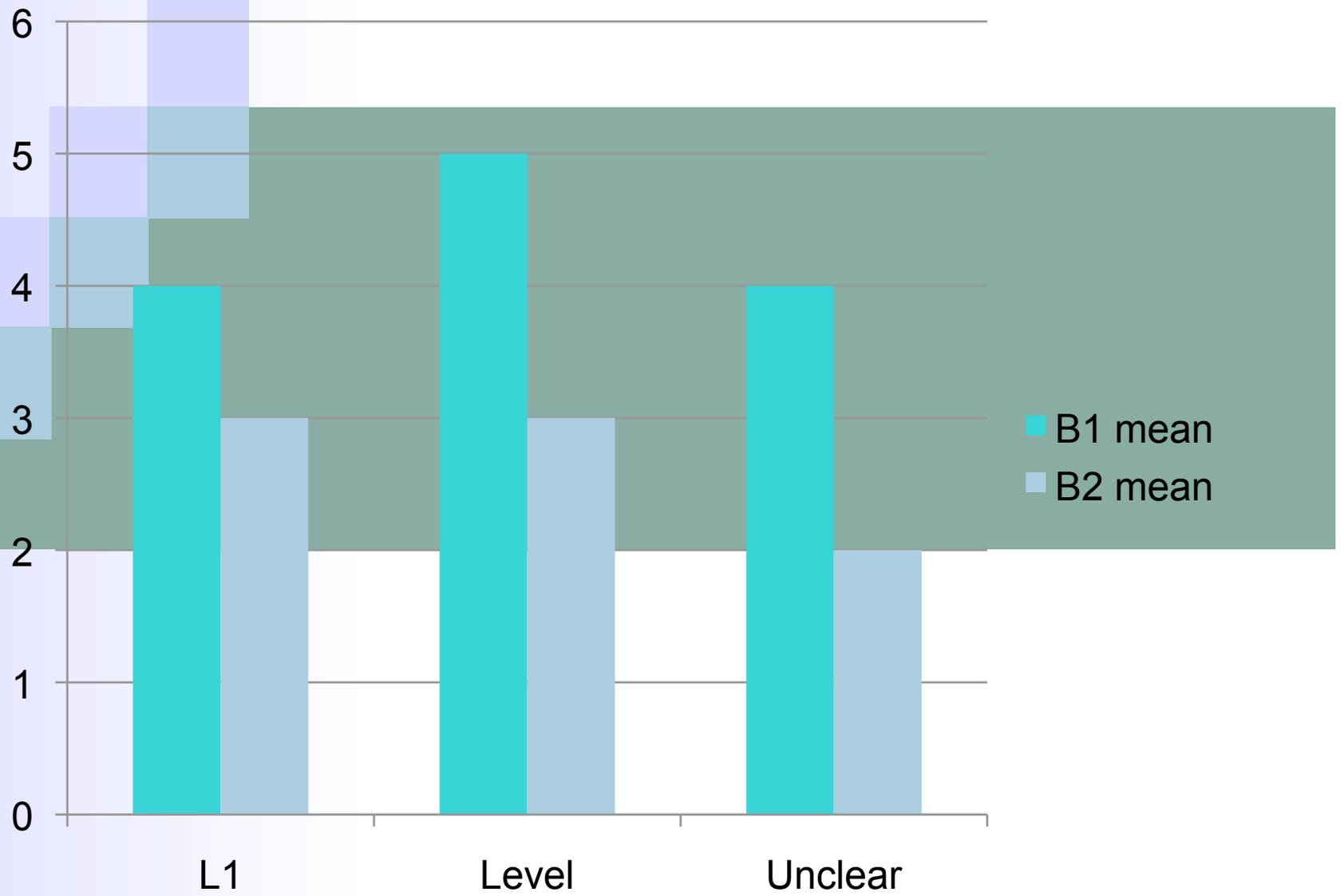
Evaluation
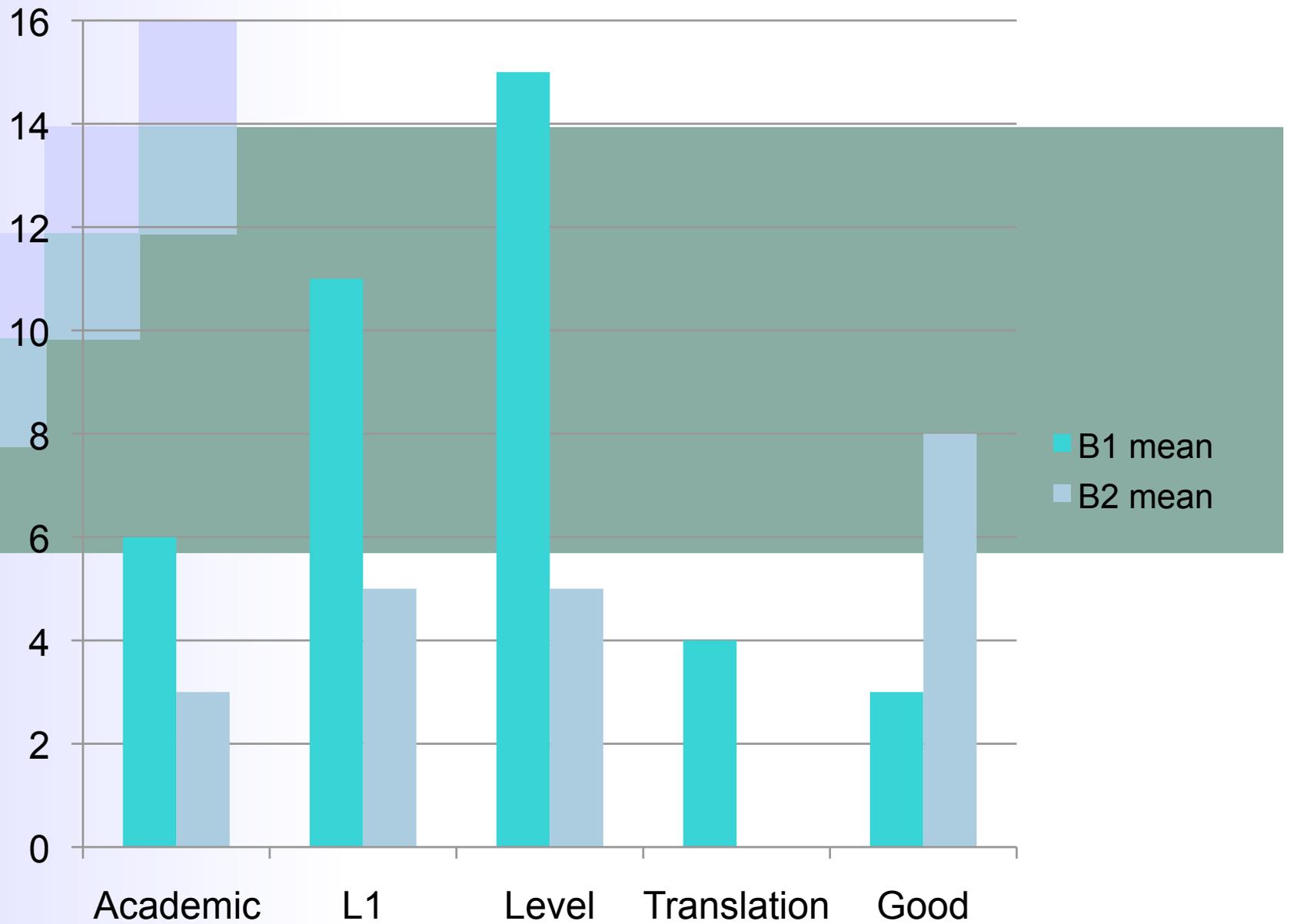
Evaluation: writing of abstracts
(based on my presentation in class)

Green = L1 / Red = level / italics = unclear / inappropriate

In the last years, several research **analyzed** the difference between native (NS) and non-native (NNS) writing for publication, using characteristic lexico-grammatical traits, text type movements and language solutions. The aim of this study is to observe these differences in writing for computer science (+ empirical / experimental, + research...), *with the view of* academic discourse competence (Spanish faculty / graduate students inform about research). The results showed that NNS used more words and grammatical constructions that NNS, and *a more restrained use of the words by NNS authors.* In conclusion, would **be advisable** to improve these language deficits implementing specific academic English courses in **research education** programs on NNS.

Evaluation

Evaluation of good choices?

*Collocations
*Text colligations)
(according to relative frequencies)
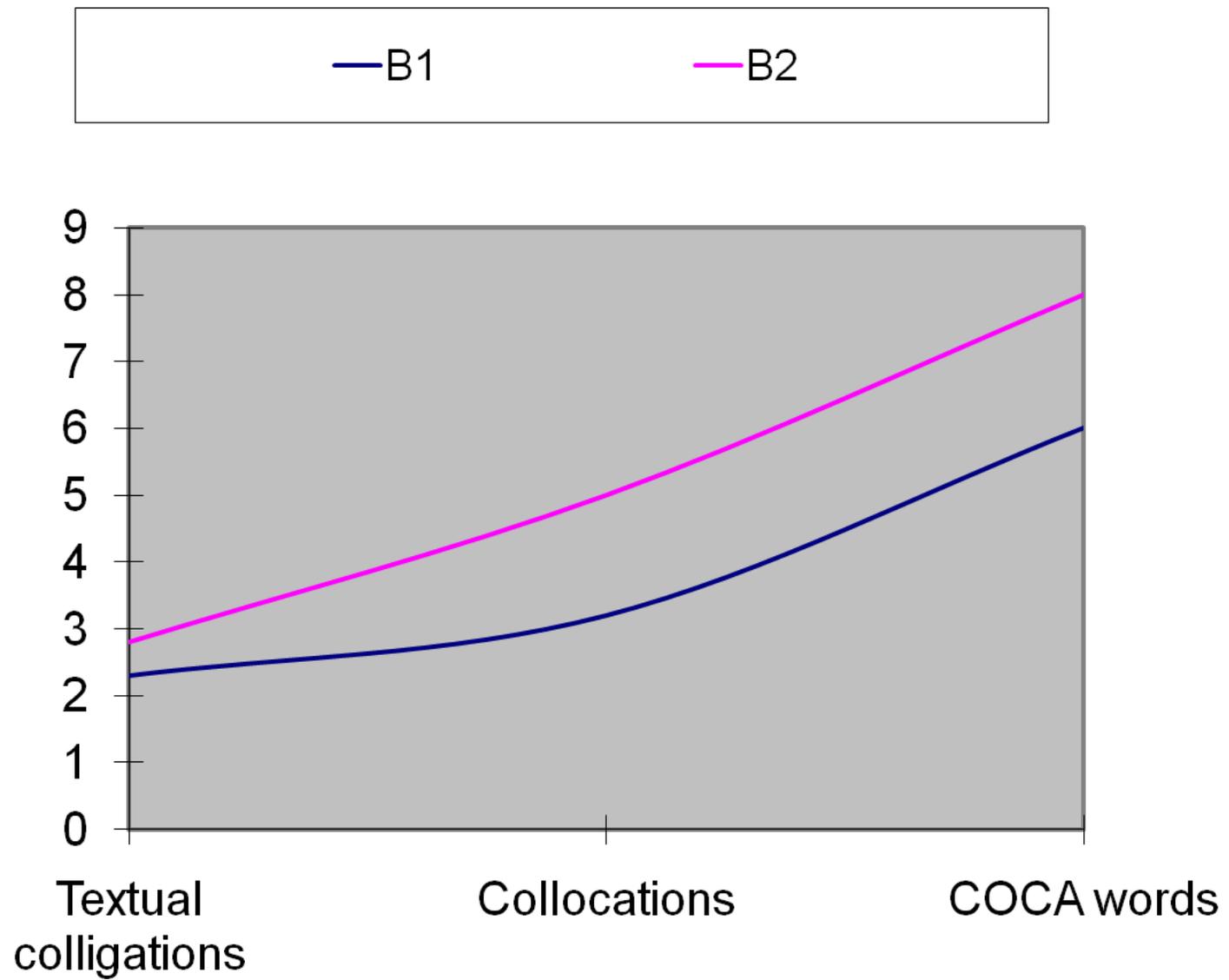+ top 100 COCA families?

e.g., from previous abstract:

observe + differences
The aim of this study is
differences, observe, important, study, research,
information, specific

Evaluation

Some concluding remarks

1. EAP / ESP > ELF (e.g., research writing, but... also, evaluative writing, e-mail interaction, ...) > Departments / administration…?

.

1. NS vs. NNS → maybe NS & NNS for more options (+ re-visiting evolving registers, i.e. contrasting corpora data, but no absolutes) Positive NNS data is received better psychologically (Mukherjee, 2009)

3. B2 > + phraseological variation = + less frequent words (advanced and Native-like: Meara, 2009) > + openness to variation

THANK YOU