Lovain-La-Neuve, 7th December

The role of collocations in text quality: towards a CEFR-graded list of collocations

ROCÍO CUBEROS & ELISA ROSADO (rcuberos@ub.edu; erosado@ub.edu)

GRUP DE RECERCA PER A L'ESTUDI DEL REPERTORI LINGÜÍSTIC (GRERLI)

UNIVERSITAT DE BARCELONA

L2 Linguistic repertoire & CEFR-Levels

Lexical correlates of linguistic proficiency

- ✓ To contribute to the empirical underpinnings of the communicative CEFR-levels (Hulstijn *et al.*, 2010)
- ✓ To determine which lexical features distinguish CEFR-levels and text quality in L2 Spanish (Cuberos, 2019)

The use of **collocations** revealed as a reliable correlate of both **text quality** and **L2 development**

Collocations in L2 curriculums

The role of collocations in language assessment

- Despite its role in L2 proficiency, collocations have not received the necessary attention in L2 curriculums or assessment material (Higueras, 2017).

- The CEFR adopts a very traditional understanding of phraseology and fails to provide a flexible definition of the term (Paquot, 2018, p. 2).

On the notion of *collocation*

"Collocations (…) are becoming prominent in our understanding of language learning and use; **however**, while the number of corpus-based LLR studies of collocations is growing, **there is still a need for a deeper understanding of factors that play a role in establishing that two words in a corpus can be considered to be collocates**"

(Glabasova, Brezina, & McEnery, 2017)

Reconciliation of approaches

Distributional approach

- Words that occur together more often than predicted by chance.

Phraseological approach

- Arbitrarily restricted combinations of lexical words.

Operational definition of *collocation*

Establishment of **specific criteria** for its identification

(Granger & Paquot, 2008; Laso, 2009)

Goal

- ✓ To present an innovative tool for the identification of collocations in L2 Spanish developed and tested in a learner corpus

- ✓ To provide a model of a CEFR-graded list of collocations in Spanish

Sample

L1 Spanish

9-11 yrs. (grade)	12-15 yrs. (junior-high)	Adults (university)	N
10	10	10	30

L2 Spanish

L2 Level	9-11 yrs. (grade)	12-15 yrs. (junior-high)	Adults (university)	N
A2	11	10	5	26
B1	7	8	11	26
B2	2	2	14	18
C1	0	0	10	10
	20	20	40	80

Developing literacy in different contexts and in different languages

PI.: R. Berman; Spanish: L. Tolchinsky

https://childes.talkbank.org/access/Spanish/GRERLI.html

Developing lingüistic repertoire in native and non-native speakers of Catalan and Spanish

PI.: J. Perera

Sample

L1 Spanish

9-11 yrs. (grade)	12-15 yrs. (junior-high)	Adults (university)	N
10	10	10	30

L2 Spanish

L2 Level	9-11 yrs. (grade)	12-15 yrs. (junior-high)	Adults (university)	N
A2	11	10	5	26
B1	7	8	11	26
B2	2	2	14	18
C1	0	0	10	10
	20	20	40	**80**

L1
Chinese
Arabic
Korean

Procedure

Developing Literacy in Different Contexts and Different Languages

Spencer Foundation, R. A. Berman, PI (Berman & Verhoeven, 2002)

- After watching a silent video about conflict situations at school, participants were asked to:
  - to <u>write</u> and to <u>tell</u> a personal experience about "a similar situation" in which they had been involved, (**narrative text**)
  - to discuss <u>orally</u> and in <u>writing</u>, the "kind of problems" that were shown in the video (**expository text**).

- Productions audiorecorded, transcribed and coded in CHAT from *CHILDES*
- Analyses performed on mean proportions over the total number of clauses in each text

TIPUS *(Test for the Identification of Phraseological Units)*

3 criteria

Step 1:   Phraseological

Step 2:   Distributional

Step 3:   Linguistic authority

A 3-step identification test!

TIPUS

Step 1: Phraseological criterion

Researcher-based identification process [Inter-rater reliability in 20% of the corpus (α = .76)]

- Collocations refer to frequent syntagmatic combinations between which a semantic link is established (Corpas, 2001).

✓ **hecho a mano**
made by hand
(hand made)

✓ **hacer trampas**
make traps
(to cheat)

✓ **emerger conflictos**
emerge conflicts
(to emerge conflicts)

✓ **defecto físico**
defect physical
(physical defect)

✓ **decir palabrotas**
say swear word
(to swear)

✓ **herida psicológica**
wound psychological
(psychological wound)

TIPUS

Step 2: Distributional criterion

Strength of the association & frequency of cooccurrence measured using *log-Dice* score (Rychlý, 2008).

- Collocations refer to words that occur together more often than predicted by chance.

Lexical association measures
- *MI* score
- *T*-score
- *log-Dice* score

Sketch Engine

➤ It operates on a scale with a fixed maximum value, which makes log-Dice **comparable across corpora** (Curran, 2003; El Maarouf & Oakes, 2015; Glabasova *et al.*, 2017)

TIPUS

Step 2: Distributional criterion

Maximum value = 14
Mean value = 10
Minimum value = 0

↓

14 – 10: Strong association
9 – 5: Medium association
4 – 0: Low association

Collocation candidates	*log-Dice*
✓ Hacer a mano (*to hand made*)	8.52
✓ Emerger conflicto (*to emerge confict*)	5.70
✓ Defecto físico (*physical defect*)	5.63
x Hacer trampas (*to cheat*)	4.79
x Herida psicológica (*psychological wound*)	2.35
x Decir palabrotas (*to swear*)	1.78

TIPUS

Práctico: *Diccionario combinatorio del español actual* (Bosque, 2006)

- Collocations are particular cases of **lexical selection**

x   hecho a mano
made by hand
(hand made)

✓   hacer trampas
make   trap
(to cheat)

x   emerger conflictos
emerge   conflicts
(to emerge conflicts)

✓   defecto físico
defect   physical
(physical defect)

✓   decir palabrotas
say   swear word
(to swear)

x   herida psicológica
wound psychological
(psychological wound)

TIPUS

Step 1: Phraseological criterion

Step 2: Distributional criterion

Step 3: Authority criterion

	Collocation candidates	log-Dice	REDES
✓	**Emerger conflicto** (*to emerge confict*)	YES	X
✓	**Hacer a mano** (*hand made*)	YES	X
✓	**Defecto físico** (*physical defect*)	YES	YES
✓	**Hacer trampas** (*to cheat*)	X	YES
✓	**Decir palabrotas** (*to swear*)	X	YES
x	Herida psicológica (*psychological wound*)	X	X

Towards a CEFR-graded list of collocations

A2

cometer un error
commit a mistake
(*to make a mistake*)

dar miedo
give fear
(*to scare*)

contar historias
tell stories
(*to tell stories*)

hacer chuletas
make chops
(*to cheat*)

sacar fotos
extract photos
(*to take photos*)

B1

dársele bien/mal
give-itself-it good/bad
(*to be good/bad at*)

hacer daño
make hurt
(*to hurt*)

darse cuenta
give-itself account
(*to realize*)

dar lo mismo
give it same
(*to be indifferent*)

consumir drogas
consume drugs
(*to take drugs*)

B2

aflorar sentimientos
appear feelings
(*to emerge feelings*)

hacer caso
make case
(*to pay attention*)

aguantar la risa
stifle the laugh
(*to stifle a laugh*)

cambiar de trabajo
change of job
(*to change jobs*)

cara de pocos amigos
face of few friends
(*grim face*)

C1

saberle mal
taste-it bad
(*to feel bad about sth*)

tener piedad
have mercy
(*to have mercy*)

ámbito educativo
field educational
(*educational sphere*)

llegar lejos
arrive far
(*to go far*)

cambiar de tema
change of subject
(*to change the subject*)

Implications of using TIPUS

✓ It captures L2 development

✓ It can be easily replicated

Operational definition of *collocation*

Establishment of **specific criteria** for its identification

- Reverse application of the identification method (v. Salido et al., 2019):
  1. Distributional criterion
  2. Phraseological criterion

Merci
Beaucoup

thank
you

Muchas
Gracias

Lovain-La-Neuve, 7th December

The role of collocations in text quality: towards a CEFR-graded list of collocations

ROCÍO CUBEROS & ELISA ROSADO (rcuberos@ub.edu; erosado@ub.edu)

GRUP DE RECERCA PER A L'ESTUDI DEL REPERTORI LINGÜÍSTIC (GRERLI)

UNIVERSITAT DE BARCELONA