

ERROR ANALYSIS IN A WRITTEN LEARNER CORPUS FROM SPANISH SPEAKERS EFL LEARNERS. A CORPUS BASED STUDY

María Victoria Pardo Rodríguez

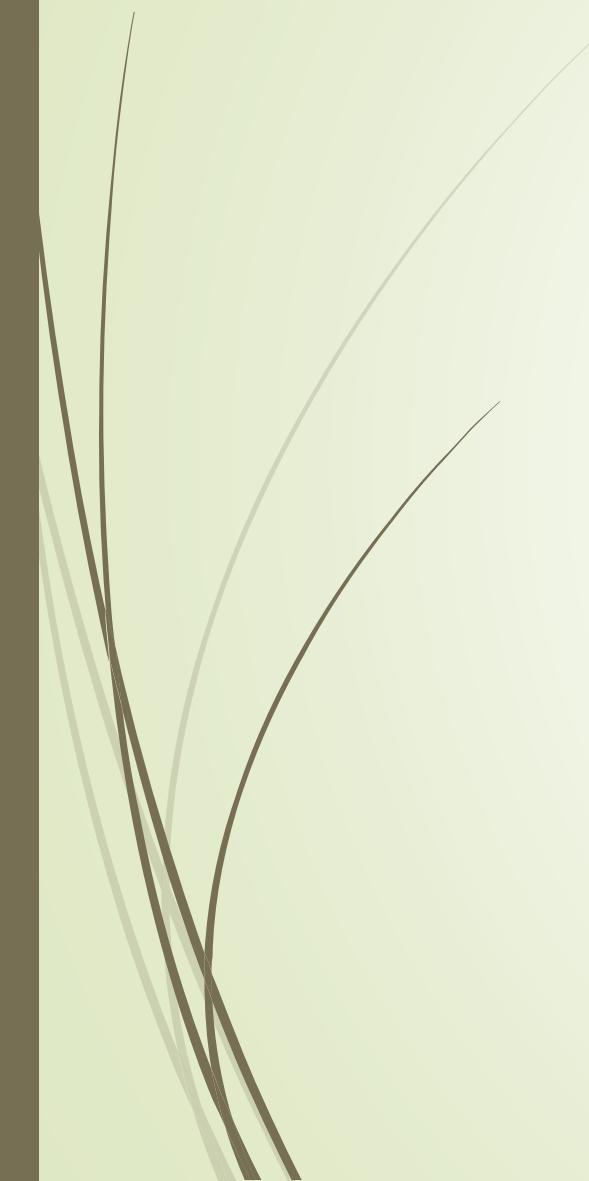
UCREL Session

Lancaster University

November 30th, 2017



Work plan

- 1. Problem summary, hypothesis, error definition.
 - 2. Compilation of the learner corpus
 - 3. Corpus' features.
 - 4. Preliminary results from pilot test including all data.
 - 5. Types of errors by category.
 - 6. Alignment of texts by type of error.
 - 7. Frequency of errors by categories.
 - 8. Types of errors compared by levels.
 - 9. Absolute and relative frequency of errors.
 - 10. CLEC Colombian Learner English Corpus.
- 



Problem summary

Problem:

- The recurrent errors in the written production of students of English as a foreign language (EFL) in Universidad del Norte from Barranquilla, Colombia
- Hypothesis to test: the input hypothesis (Krashen, 1982). Language is acquired by receiving “comprehensible input” (CI) slightly above the current level of competence...grammar is automatically acquired if there is enough CI
- How proficiency changes from level to level
- **Error**, defined by James (1998) as “...an instance of language that is unintentionally deviant and is not self-correctible by its author.” (P. 78).

Compilation of the learner corpus I

Third semester:

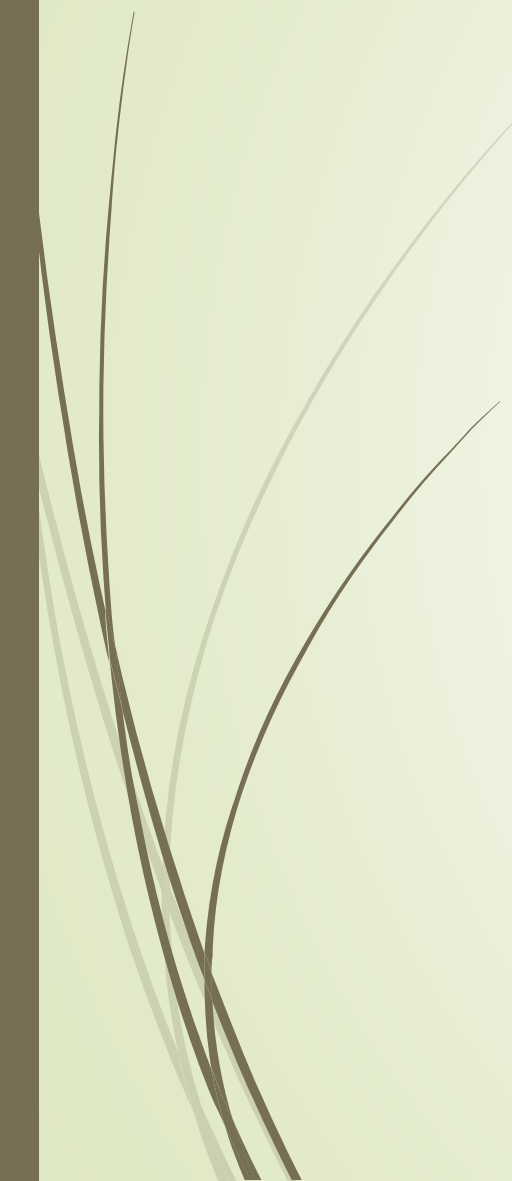
- Arrangement of student's work in different files. In total 518 students authorized the use of their data for research purposes.
- Louvain university was contacted. We bought an error tagger for EFL errors.

Fourth semester:

- **Handwritten** assignments were **transcribed** into digital files, saved as TXT files and were assigned special codes to make them traceable.
- Manual error tagging starts.



Compilation of the learner corpus II

- The files were error tagged and put together by levels.
 - Papers were aligned according to the type of error in WordSmith (WS).
 - The first findings were organized in Excel sheets and errors were filtered according to each category
- 



Compilation of the learner corpus III

- External review started to check consistency, and correct tagging. (EFL expert)
- First pilot findings were presented in the First Corpus and Computational Linguistics International Congress.
(Caro y Cuervo Institute. Bogotá, Colombia)

Example from a written file into digital file

L3D3

NOMBRE Ariana Rodriguez H. FECHA 17th sept 2015
INSTITUCION Unimorte PROFESOR
MATERIA CURSO NOTA 4

Final draft

✓ In my opinion, ~~some~~ sometimes tourists are dangerous for ~~the~~ environment. First, ~~tourists~~ tourists aren't care in place visited. For example, when tourists visited the coast ~~don't~~ ^{they} don't clean this place. Second, the pollution ^{prep} on the ecotourism ^{su} are more than another places. For example, the rivers, see, Paramount. Finally, tourists ^{do think it is} not important ^{to} protected the place ^{key} visited. This is terrible! For example, natural places and animals are on extinction. In conclusion, tourists in a future take

Ana Rodriguez H. - L3D3

Final Work:

In my opinion, sometimes tourists are dangerous for environment. First, tourists aren't care in place visited. For example, when tourists visited the coast don't clean this place. Second, the pollution on the ecotourism are more than another places. for example, the rivers, see, paramount. Finally, tourists not important protected the place visited. This is terrible! For example, natural places and animals are on extinction. In conclusion, tourists in a future take care more on ecotourism. Is very important save the natural places because are unique.



Errors by categories (Louvain University)

- **Formal errors F**
- **Grammatical errors, i.e. errors that break general rules of English grammar G**
- **Lexico-grammar errors, i.e. errors where the morpho-syntactic properties of a word have been violated X (XADJ, XVPR...)**
- **Lexical errors, i.e. errors involving the semantic properties of single words and phrases LS**
- **Word Redundant, Word Missing and Word Order errors WO, WR**
- **Punctuation errors QM, QR**
- **Style errors SI, SU**
- **Infelicities Z**

Examples of some errors tagged


- **37** another reason is that they (Z) wanna \$ want to\$ show a
- **113** could be a good way to try (XVPR) 0 \$to\$ survive with canc
- **484** But in contrast, there are too (WRS) too\$0\$ (XNUC) much \$many\$ people
- **6536** tor examines our body, he can (GWC) diagnostic \$diagnose\$ US
- **8431** are not honest. The product (GVAUX) 0 \$does\$ not see
- **11041** ... emotions. For example, when (GA) the \$0\$ people see commercials
- **13426** so for example Shakira is a Colombian (FS) celebritie \$celebrity\$

Digital file becomes TXT file and is error tagged

18/ sept/ 2015

I think that the ecotourism isn't good for ou planet for differents reasons. The first reason is that this practice produce contamination because the people don't take care and they don't clean the place. For example, when I went to the Tayrona Park and I saw this place isn't clean because they throught waste in all ecosystem. The second reason is that the ecotourism affect to animal's home because this place is habitted for many people and it does that the animals go out. Finally the ecotourism changes the ecosystem and it affect the climate because it can change the temperature. In conclusion the ecotourism is evil and dangerous for the planet because the people isn't friendly with the nature. Is a problem that affect to all world. Is necessary thinking a solution for this problem right now.

I think that (GA) the \$0\$ ecotourism isn't good for (FS) ou \$our\$ planet for (GADJN) differents\$different\$ reasons. The first reason is that this practice (GVN) produce \$produces\$ contamination because (GA) the \$0\$ people don't take care and they don't clean the place. For example, when I went to the Tayrona Park and I saw this place isn't clean because they (LS) throught \$throw\$ waste in all (GA) 0 \$the\$ (FS) ecosystem\$ecosystem\$. The second reason is that (GA) the \$0\$ ecotourism (GVN) affect \$affects\$ (LP) to animal's home \$animal's environment\$ because (LP) this place is (FS) habitted \$inhabited\$ (LS) for\$by\$ many people (LP) and it does that the\$so\$ animals (LP)go out \$leave\$ \$when people arrive, they displace animals\$. Finally (QM)0 \$,\$ (GA)the \$0\$ ecotourism changes the ecosystem and it (GVN) affect \$affects\$ the climate because it can change the temperature. In conclusion (GA) the \$0\$ ecotourism is evil and dangerous for the planet because (GA) the\$0\$ people (GVN) isn't \$aren't\$ friendly with (GA) the \$0\$ nature. (GPP) 0 \$It\$ Is a problem that (GVN) affect \$affects\$ (WRS) to \$0\$ all world. (GPP) 0 \$it\$ (FS) Is \$is\$ necessary (GWC) thinking \$to think\$ (WM) 0 \$about\$ a solution for this problem right now.



Corpus' features

- Total of words: 151.708
- Range of words per paper 50 – 1.300
- Median of words per paper: 292
- Vocabulary richness (density): 8.112 (use of content words)
- Number of sentences in all corpus: 5.947

Alignment of texts by type of error

file	Edit	View	Compute	Settings	Windows	Help
1	Concordance					Set Tag V
	\$not having done\$	more things in	your (GWC)	live \$life\$.	Your family will	
	sports are determined by the way	you (GWC)	practiced \$practice\$	them		
	is time to start to change your life if	you (GWC)	considered \$consider\$	that		
	sports, we can implicate this	with (GWC)	a dead \$death\$,	but we need		
	could affect people (GWC)whom	\$who\$ (GWC)	life \$live\$	in those ways.		
	think that (GA) the \$0\$	one way to (GWC)	death \$die\$ (GWC)	easy		
	of living can (GVN) causes	\$cause\$ the (GWC)	feel \$feeling\$	that everything		
		THE (GWC)	SILENCE \$silent\$	RISKS OF		
	organs. The sedentary life style is	the (GWC)	originating \$origin\$	of		
0	way, the sedentary life can produce	the (GWC)	die \$death\$	too if we (Z) can't		
1	sedentary life are risky, but both in	the (GWC)	reality \$real\$	life are		
2	\$carry on\$ to the same destiny. The	(GWC)	different \$difference\$	is that in		
3	there are some people whom think	that (GWC)	sedentary \$sedentarism\$	is a		
4	and aggressive political leader	doesn't (GWC)	intervened \$intervene\$	and		
5	is the sedentary life (WM) \$style\$.	(GWC)	Have \$having\$	a (GWC) relax		
6	face risks even if they live a	sedentary (GWC)	live \$life\$,	but these risks can		
7	that it (GVN) require	\$requires\$, (GWC)	too \$to\$	have (LSF) distinct		
8	people could play extreme sports	or (GWC)	life \$live\$ (LP)	as sedentary		
9	like eating junk food all the time,	or (GWC)	stay \$staying\$	on TV for hours		
0	lifestyle can also have a lot of risks	like, (GWC)	loose \$loosing\$	all the energy		
1	\$to\$ find a form to keep us and our	kids (GWC)	save \$safe\$. Today the kids		
2	(GVN) want \$wants\$ (LS) 0 \$to\$	keep (GWC)	save \$safe\$ (LP)	want keep		
3	exists \$exist\$ a lot of ways to	keep (GWC)	save \$safe\$	and reduce this		
4	state the only action that you do	is (GWC)	eat \$eating\$	bad food), being		
5	the HPV effects in the health, as	in (GWC)	deficit \$deficient\$	information		

First pilot testing analysis: Total of errors tagged: 14.531

The screenshot displays a Microsoft Word document titled "Documento1 - Word" with the "HERRAMIENTAS DE IMAGEN" ribbon active. The main content is a concordance table with the following columns: N, Concordance, Set, Tag, Word #, Sen, Para, Para, lead, lead, Sec, Sec, File, and %. The table lists 60 rows of text with corresponding error tags and statistics. A red circle highlights the total number of errors, 14,531, in the bottom left corner of the table.

N	Concordance	Set	Tag	Word #	Sen	Para	Para	lead	lead	Sec	Sec	File	%
32	this manner (QM) 0 S,\$ (GPU) 0 \$they\$	(Z)	will not \$won't\$ be involved in danger	562	225%	034%				034%	Steve23SecondC		84%
33	\$flashfloods\$ some people lose their	(Z)	stuff \$belongings\$, (GNN)	742	1213%	079%				079%	SecondCorrection		80%
34	, fatal. On the other hand, (GA) the \$0\$	(Z)	90 \$nety\$ percent of those with	315	1329%	047%				047%	Steve23SecondC		45%
35	(GNN) flood \$floods\$ Barranquilla is the	(Z)	4th \$fourth\$ business city in the	9	027%	0 1%				0 1%	SecondCorrection		1%
36		(Z)	The risks of living outside of Colombia	1	0 6%	0 0%				0 0%	SecondCorrection		0%
37	method \$methodology\$ to attend all kind	(XVPR)	\$to \$of\$ students and increase	236	932%	051%				051%	ave4SecondC.tx		52%
38	flame \$blaze\$. Finally, tourists don't	(XVPR)	care \$care\$ about\$ the place that	200	836%	031%				031%	vid Smart Grupo		81%
39	people do not change, and they only	(XVPR)	wait 0 \$wait\$ for\$ the opportunity	37	154%	016%				016%	PDF 47 Toti level		17%
40	(GWC) depends \$depending\$	(XVPR)	\$of \$on\$ the kind of (GNN) tourist	24	075%	012%				012%	3Vides22PDF4		13%
41	a danger to the environment. It is	(XVPR)	\$depending\$ of \$depending\$ on\$ the	15	127%	0 8%				0 8%	MARILLOPDF2		9%
42	bad or happy, (FS) according \$according\$	(XVPR)	\$according\$ with \$according\$ to\$	102	359%	050%				050%	ardo Rojas 5613		62%
43	\$usually\$ appears (WRM) most on \$0\$	(XVPR)	\$in\$ the last months of the year.	222	615%	034%				034%	ave6SecondC.tx		32%
44	a conviction (GVN) depend \$depend\$	(XVPR)	\$of \$on\$ (GA)0 \$the\$ crime (LP) it	46	147%	0 8%				0 8%	Steve13SecondC		8%
45	0 \$one\$ another\$, (GPP) \$they\$ don't	(XVPR)	care 0 \$care\$ about\$ the age.	151	355%	040%				040%	PDF 46 Toti level		38%
46	ago people (GVT) search \$searched\$	(XVPR)	search 0\$searched\$ for\$	142	518%	070%				070%	R L 4 NRC 5642		70%
47	(QM)0 \$,\$ the inhabitants are affected	(XVPR)	\$by\$ mobility to their homes.	434	1530%	054%				054%	SecondCorrection		65%
48	the deaths and accidents recklessly dur	(XVPR)	\$to\$ flashfloods, the Water	315	840%	048%				048%	ave6SecondC.tx		45%
49	, (GPP) 0 \$he\$ (GVAUX) don't \$doesn't\$	(XVPR)	care 0 \$care\$ about\$ the age	57	238%	015%				015%	PDF 46 Toti level		15%
50	in the earth \$on\$ the ground\$ and look	(XVPR)	\$at\$ the scenic landscape.	251	1250%	032%				032%	L34996-1PDF11		92%
51	(GA) The \$0\$ young (WM)0 \$people\$	(XVPR)	think in \$think\$ about\$ (LPF) a	66	344%	032%				032%	R 4P44 PDF 39		32%
52	. Your mother or your father will no	(XVPR)	look your things \$looks\$ at your	91	456%	051%				051%	vid Smart Grupo		51%
53	inside your own community regardles	(XVPR)	\$of\$ their special needs.	314	1037%	058%				058%	3 Secondchecke		59%
54	(QC) , \$,\$ we (GWC) listening \$listens\$	(XVPR)	\$to\$ a lot of (LSF) notice	610	1037%	055%				055%	SecondCorrection		66%
55	this \$that\$ (FS)\$people\$ \$people\$ don't	(XVPR)	think in \$think\$ about\$ the (LS)	213	875%	031%				031%	PDF 47 VERDE		81%
56	live in\$0\$ the real life where no one	(XVPR)	depends by \$depends\$ on\$ his	204	733%	073%				073%	DF 81 BLANCO		70%
57	psychology□ the institutions depend	(XVPR)	\$of \$on\$ the profile or case, each	326	757%	053%				053%	Steve13SecondC		54%
58	help people with this disease; it depends	(XVPR)	\$of \$on\$ (GA)0 \$the\$ type of	71	256%	0 7%				0 7%	Steve14SecondC		7%
59	any \$some\$ people might not agree	(XVPR)	\$with\$ my opinion about gender	622	2538%	052%				052%	level 8-2 Second		53%
60	criminals and make people absta	(XVPR)	\$of \$from\$ committing crimes.	122	436%	026%				026%	SecondCorrectio		25%

14,531 Set

Types of errors by categories I

Congreso_CaroyCuervo_2017_VP-GQ - PowerPoint

TOTAL ERRORES TODOS NIVELES_GQ [Modo de compatibilidad] - Excel

	A	B	C	D	E	F	G
5							
6	Clase	Valor	Frecuencia	Frecuencia relativa	Frecuencia Acumulada		
7	1	FM	8	0,0006	8		
8	2	FS	1917	0,1319	1925		
9	3	FSR	6	0,0004	1931	13,29	
10	4	GA	1713	0,1179	3644		
11	5	GADJCS	68	0,0047	3712		
12	6	GADJN	113	0,0078	3825		
13	7	GADJO	13	0,0009	3838		
14	8	GADVO	33	0,0023	3871		
15	9	GDD	64	0,0044	3935		
16	10	GDI	64	0,0044	3999		
17	11	GDO	116	0,008	4115		
18	12	GDT	31	0,0021	4146		
19	13	GNC	100	0,0069	4246		
20	14	GNN	811	0,0558	5057		
21	15	GPD	82	0,0056	5139		
22	16	GPE	82	0,0056	5221		

Búsqueda en Windows

Vínculos

2:21 p. m.
11/05/2017

Types of errors by categories II

Congreso_CaroyCuervo_2017_VP-GQ - PowerPoint

HERRAMIENTAS DE IMAGEN

TOTAL ERRORES TODOS NIVELES_GQ [Modo de compatibilidad] - Excel

	B	C	D	E	F
22	GPF	9	0,0006	5148	
23	GPI	34	0,0023	5182	
24	GPO	17	0,0012	5199	
25	GPP	551	0,0379	5750	
26	GPR	112	0,0077	5862	
27	GPU	112	0,0077	5974	
28	GVAUX	373	0,0257	6347	
29	GVM	65	0,0045	6412	
30	GVN	656	0,0451	7068	
31	GVNF	25	0,0017	7093	
32	GVT	240	0,0165	7333	
33	GVV	51	0,0035	7384	
34	GWC	739	0,0509	8123	42,6
35	LCC	121	0,0083	8244	
36	LCLC	55	0,0038	8299	

Búsqueda en Windows

Vínculos

2:23 p. m.
11/05/2017

Types of errors by categories III

Congreso_CaroyCuervo_2017_VP-GQ - PowerPoint

TOTAL ERRORES TODOS NIVELES_GQ [Modo de compatibilidad] - Excel

	A	B	C	D	E	F	G	H	I
34	28	GWC	739	0,0509	8123	42,6			
35	29	LCC	121	0,0083	8244				
36	30	LCLC	55	0,0038	8299				
37	31	LCLS	3	0,0002	8302				
38	32	LCS	16	0,0011	8318				
39	33	LP	864	0,0595	9182				
40	34	LPF	167	0,0115	9349				
41	35	LS	1255	0,0864	10604				
42	36	LSF	181	0,0125	10785	18,33			
43	37	QC	227	0,0156	11012				
44	38	QL	17	0,0012	11029				
45	39	QM	611	0,042	11640				
46	40	QR	91	0,0063	11731	6,51			
47	41	SI	19	0,0013	11750				
48	42	SU	500	0,0344	12250	3,57			
49	43	WM	645	0,0444	12895				
50	44	WO	328	0,0226	13223				
51	45	WRM	347	0,0239	13570				
52	46	WRS	668	0,046	14238	13,69			
53	47	XADJPR	7	0,0005	14245				
54	48	XNCO	1	0,0001	14246				

Búsqueda en Windows

Vínculos

2:24 p. m. 11/05/2017

Types of errors by categories IV

Congreso_CaroyCuervo_2017_VP-GQ - PowerPoint

TOTAL ERRORES TODOS NIVELES_GQ [Modo de compatibilidad] - Excel

	A	B	C	D	E	F	G	H	I
40	34	LPF	167	0,0115	9349				
41	35	LS	1255	0,0864	10604				
42	36	LSF	181	0,0125	10785	18,33			
43	37	QC	227	0,0156	11012				
44	38	QL	17	0,0012	11029				
45	39	QM	611	0,042	11640				
46	40	QR	91	0,0063	11731	6,51			
47	41	SI	19	0,0013	11750				
48	42	SU	500	0,0344	12250	3,57			
49	43	WM	645	0,0444	12895				
50	44	WO	328	0,0226	13223				
51	45	WRM	347	0,0239	13570				
52	46	WRS	668	0,046	14238	13,69			
53	47	XADJPR	7	0,0005	14245				
54	48	XNCO	1	0,0001	14246				
55	49	XNPR	3	0,0002	14249				
56	50	XNUC	43	0,003	14292				
57	51	XPRCO	89	0,0061	14381				
58	52	XVCO	7	0,0005	14388				
59	53	XVPR	107	0,0074	14495	1,78			
60	54	Z	36	0,0025	14531	0,2			

Búsqueda en Windows

Vínculos

2:25 p. m. 11/05/2017

Frequency of errors by categories

Cat. error	Percent.	Frequency
Grammar	42,6	6192
Lexis	18,33	2662
W	13,69	1988
F	13,29	1931
Q	6,51	946
S	3,57	519
X (LG)	1,78	257
Z	0,2	36
Totals	100%	14531

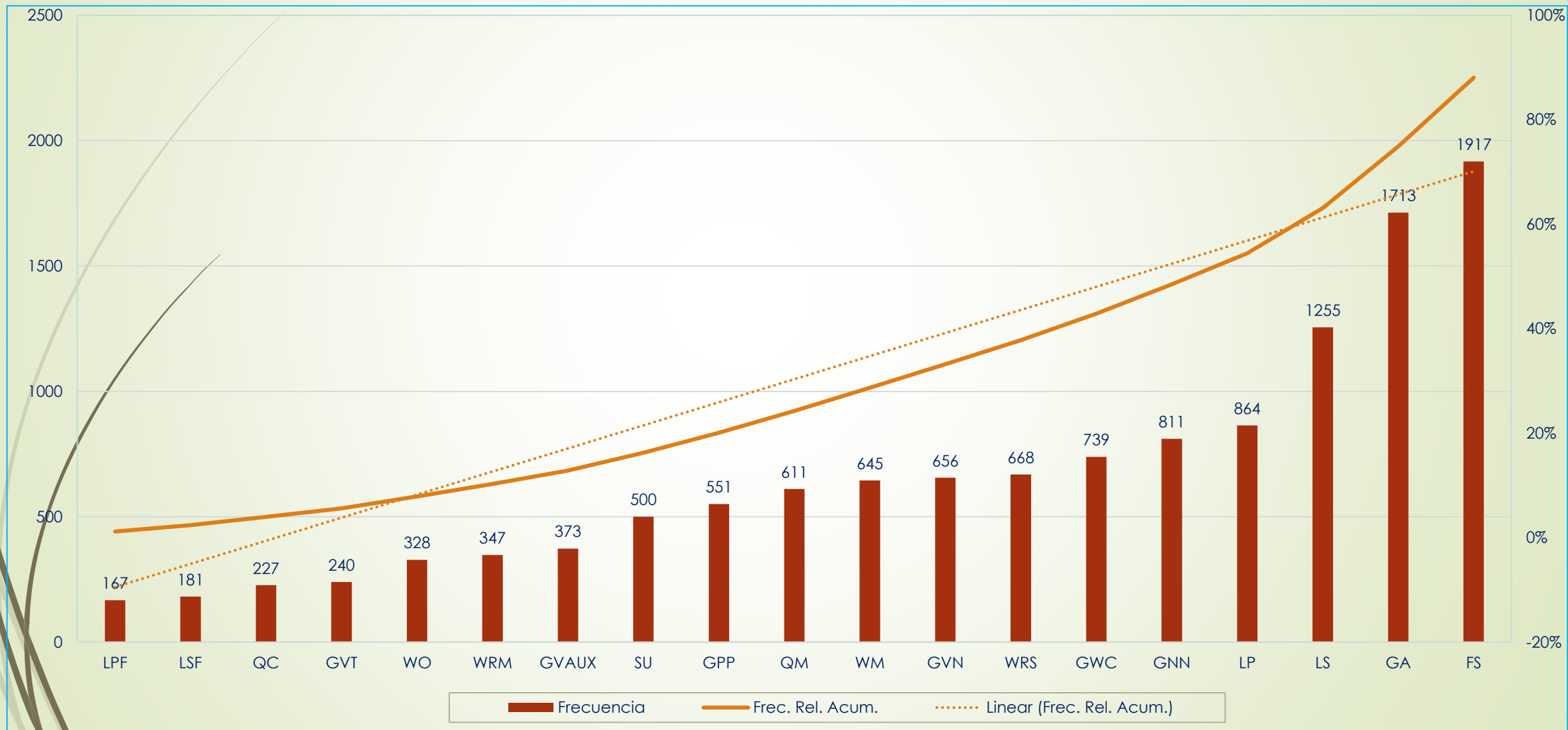
Comparative chart by type of errors in different levels I

A1			A1.2			B1			B1.3 & B2		
Error	Frequency	Percentage	Error	Frequency	Percentage	Error	Frequency	Percentage	Error	Frequency	Percentage
FS	1.040	18,35%	FS	529	16,44%	FS	119	20,70%	LS	579	11,42%
GA	836	14,75%	GA	361	11,22%	GA	90	15,65%	GA	426	8,40%
LS	441	7,78%	QM	205	6,37%	GNN	44	7,65%	GWC	355	7,00%
GNN	374	6,60%	LS	199	6,18%	LS	36	6,26%	WRS	347	6,84%
LP	349	6,16%	LP	185	5,75%	SU	35	6,09%	GNN	308	6,07%
WM	312	5,50%	SU	178	5,53%	GVAUX	27	4,70%	LP	308	6,07%
GVN	277	4,89%	GWC	170	5,28%	LP	22	3,83%	QM	242	4,77%
WRS	200	3,53%	WM	151	4,69%	GVN	20	3,48%	FS	229	4,52%
GWC	195	3,44%	GPP	150	4,66%	QM	20	3,48%	GVN	221	4,36%
GPP	179	3,16%	GVN	138	4,29%	WRS	20	3,48%	GPP	203	4,00%

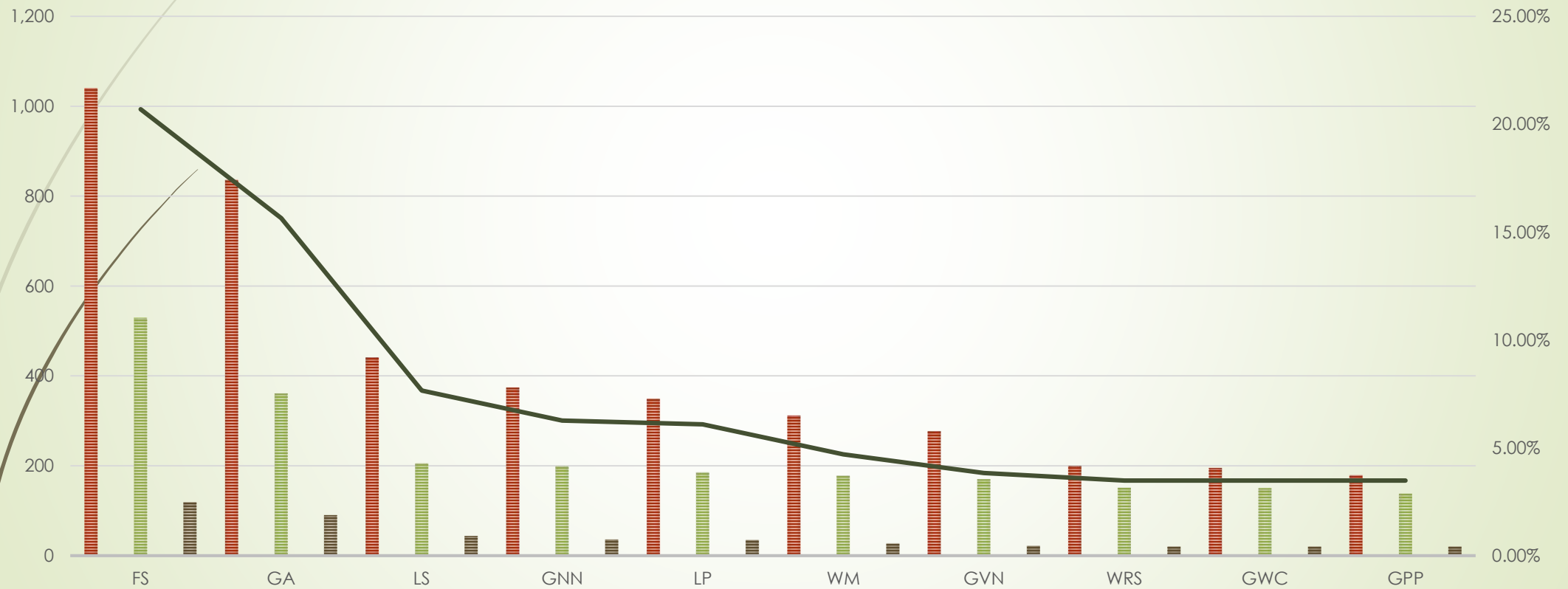
Absolute and relative frequency of errors chart.

Error	A. Frequency	Relt. Freq. Acum.	Relative Freq.
LPF	167	1%	0,0115
LSF	181	2%	0,0125
QC	227	4%	0,0156
GVT	240	6%	0,0165
WO	328	8%	0,0226
WRM	347	10%	0,0239
GVAUX	373	13%	0,0257
SU	500	16%	0,0344
GPP	551	20%	0,0379
QM	611	24%	0,042
WM	645	29%	0,0444
GVN	656	33%	0,0451
WRS	668	38%	0,046
GWC	739	43%	0,0509
GNN	811	48%	0,0558
LP	864	54%	0,0595
LS	1255	63%	0,0864
GA	1713	75%	0,1179
FS	1917	88%	0,1319
Totales	12793	88,931	88,05

Absolute and relative frequency of errors table



Trend of the same error in three different levels A1,A2,B1

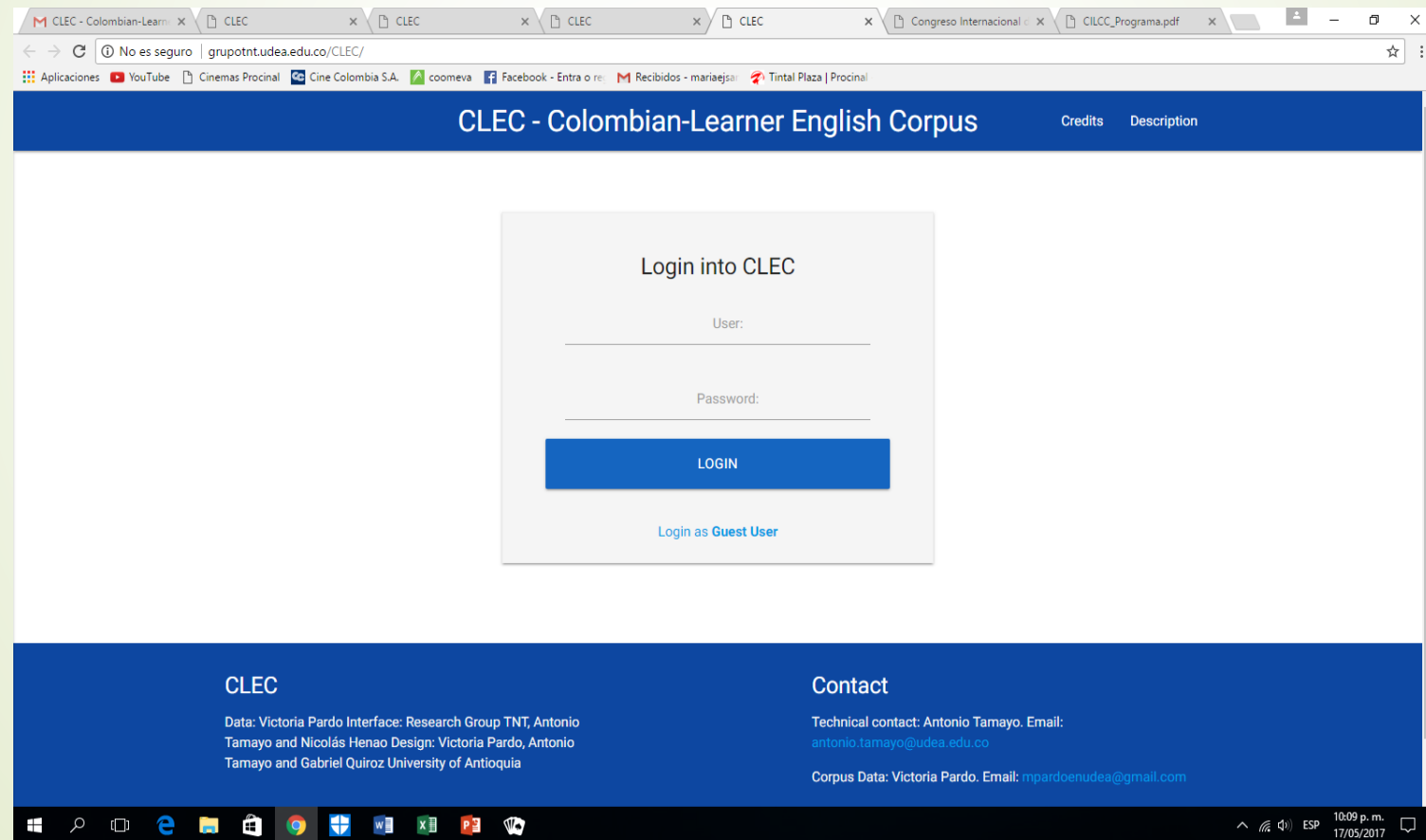


CLEC - Colombian-Learner English Corpus

<http://grupotnt.udea.edu.co/CLEC/>

<http://grupotnt.udea.edu.co/CLEC/description/index.htm>

<http://grupotnt.udea.edu.co/CLEC/credits/index.htm>



The screenshot shows a web browser window displaying the CLEC website. The browser's address bar shows the URL <http://grupotnt.udea.edu.co/CLEC/>. The website has a blue header with the title "CLEC - Colombian-Learner English Corpus" and navigation links for "Credits" and "Description". The main content area features a "Login into CLEC" form with fields for "User:" and "Password:", a blue "LOGIN" button, and a link for "Login as Guest User". The footer is also blue and contains information about the project, including the names of the data and interface designers, and contact details for the technical and corpus data contacts.

CLEC - Colombian-Learner English Corpus

Credits Description

Login into CLEC

User:

Password:

LOGIN

Login as Guest User

CLEC

Data: Victoria Pardo Interface: Research Group TNT, Antonio Tamayo and Nicolás Henao Design: Victoria Pardo, Antonio Tamayo and Gabriel Quiroz University of Antioquia

Contact

Technical contact: Antonio Tamayo. Email: antonio.tamayo@udea.edu.co

Corpus Data: Victoria Pardo. Email: mpardo@udea.edu.co

10:09 p. m.
17/05/2017

CLEC - Colombian-Learner English Corpus

Credits

This corpus has been built under the Doctoral thesis entitled "Problemas en la producción escrita del proceso de aprendizaje del inglés como lengua extranjera: un estudio de corpus en estudiantes universitarios de inglés de la costa norte de Colombia" by María Victoria Pardo, under the supervision of Dr. Gabriel Quiroz in the Doctoral Program on Linguistics at University of Antioquia.

Collected Data, correction, tagging and analysis have been done by María Victoria Pardo.

Tagging was done using the Université Catholique de Lovaine error editor software.

Corpus interface was designed by Victoria Pardo, Antonio Tamayo and Gabriel Quiroz.

Computer engineering programming design was done by Antonio Tamayo and Nicolás Henao.

Data collection was done at Universidad del Norte in Barranquilla, Colombia. Collected during the second semester 2015 from students registered in the English program from different careers. Thanks to Universidad del Norte.

To cite the obtained data, please refer as:

CLEC (Online) Medellín: Victoria Pardo and TNT Research Group, Escuela de Idiomas- Universidad de Antioquia. (On Day/Month/ Year)
(<http://grupotnt.udea.edu.co/CLEC>)

CLEC

Data: Victoria Pardo Interface: Research Group TNT, Antonio Tamayo and Nicolás Henao Design: Victoria Pardo, Antonio Tamayo and Gabriel Quiroz University of Antioquia




CLEC - Colombian-Learner English Corpus

Description

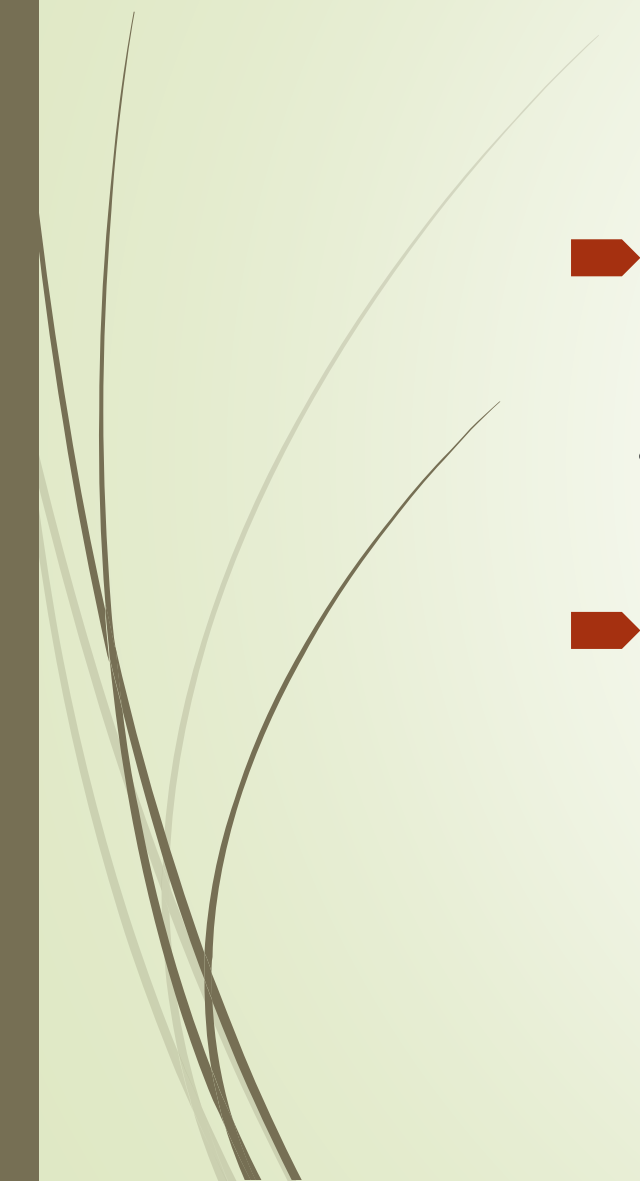
CLEC (Colombian Learner English Corpus) is a Corpus of Learner of English as a foreign language. The corpus is made up of 600 compositions of argumentative texts from English learners of basic, intermediate and upper intermediate levels. The total number of running words is approx. 200.000. The Corpus was tagged using error tagging UCLE software from the Université de Lovaine. This web app is useful for English teachers in Colombia to know statistics on most common errors of Colombian Learners and find examples. Thank you very much to those students who gave their permission to use data for the doctoral research and the Universidad del Norte for giving special attention and permission to carry out the project. Data is used anonymously to protect students' identity. Data is used solely for research purposes. Feedback on data interface, or general comments can be emailed to Victoria Pardo, Gabriel Quiroz or Antonio Tamayo.

CLEC

Data: Victoria Pardo Interface: Research Group
TNT, Antonio Tamayo and Nicolás Henao
Design: Victoria Pardo, Antonio Tamayo and
Gabriel Quiroz University of Antioquia



What's next?

- Further analysis on how students develop and progress in their interlanguage level.
 - Develop a friendlier error tagger for learner corpora.
- 



THANK YOU

Bibliografía

- ▶ Corder, P. (1988). *Error Analysis and Interlanguage*. Oxford: Oxford. [Consultado el 7 de mayo de 2017].
- ▶ Dargneaux, E., Dennes, S., Granger, S., Meunier, F., Neff, J., & Thewissen, J. (2005). *Error Tagging Manual Version 1.2*. (1st ed., pp. 23-28). Université Catholique de Louvain: Centre for English Corpus Linguistics.
- ▶ Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- ▶ Hymes, D.H. (1972) "On Communicative Competence" En: J.B. Pride and J. Holmes (eds) *Sociolinguistics. Selected Readings*. Harmondsworth: Penguin, pp. 269-293.(Part 2) Disponible en: <http://wwwhomes.uni-bielefeld.de/sgramley/Hymes-2.pdf> (consultado el día 16 de marzo de 2016).
- ▶ Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *Revista CALICO* 20(3), 465–480. URL <http://purl.org/calico/Granger03.pdf> (consultada agosto 07, 2016).
- ▶ Krashen, Stephen (2014). "Teorías de la Adquisición de una Segunda Lengua. Teoría de Krashen", sitio web de Google, [en línea]. Disponible en: <https://sites.google.com/site/adquisiciondeunasegundalengua/teorias> [consultado el día 15 de agosto de 2014].