La linguistique de corpus à l’appui des synthèses de recherche

alex.boulton@univ-lorraine.fr

ATILF, 22 mars 2024

Journée thématique et transversale :

Linguistique de corpus à la croisée de questionnements théoriques, méthodologiques et empiriques

Apprentissage sur corpus (ASC)

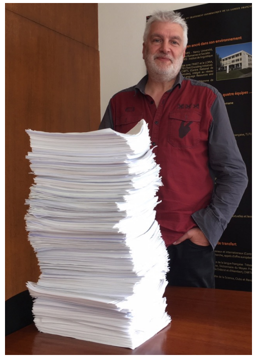
aka ‘data-driven learning’ (DDL)

> “the attempt to cut out the middleman as far as possible and… give the learner direct access to the data” (Johns, 1990, p.18)

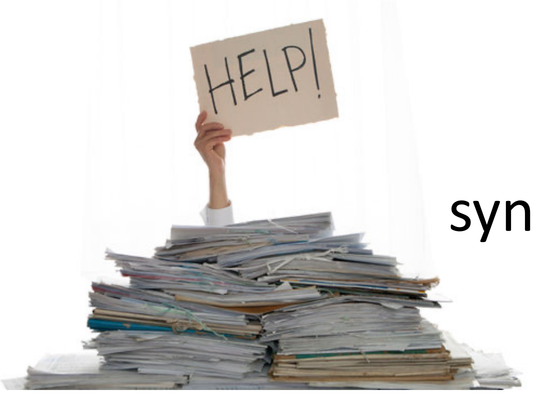
> “using the tools and techniques of corpus linguistics for pedagogical purposes” (Gilquin & Granger, 2010, p.359) ⇨ ±directly for L2

My empirical DDL collection:

2007 = 39
2012 = 116
2017 = 210
2019 = 351
2021 = 489…



now 777!

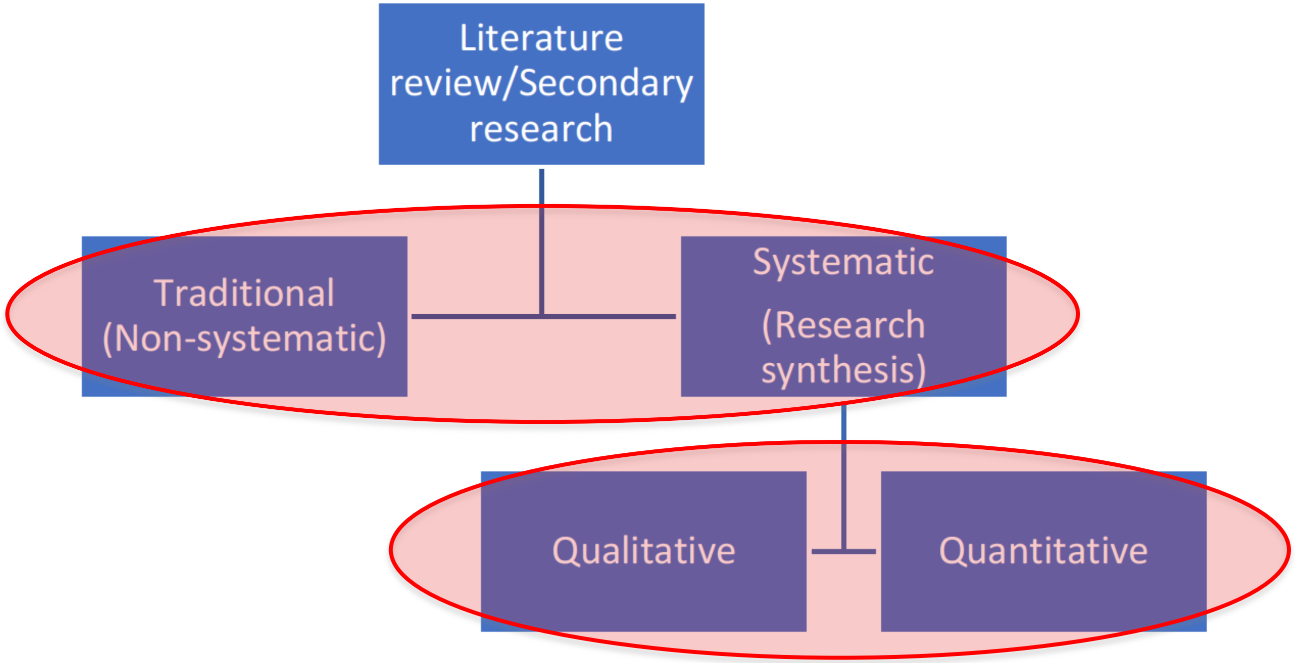


syntheses…

Research syntheses

A typology of secondary research in applied linguistics

(Chong & Plonsky, advance access)



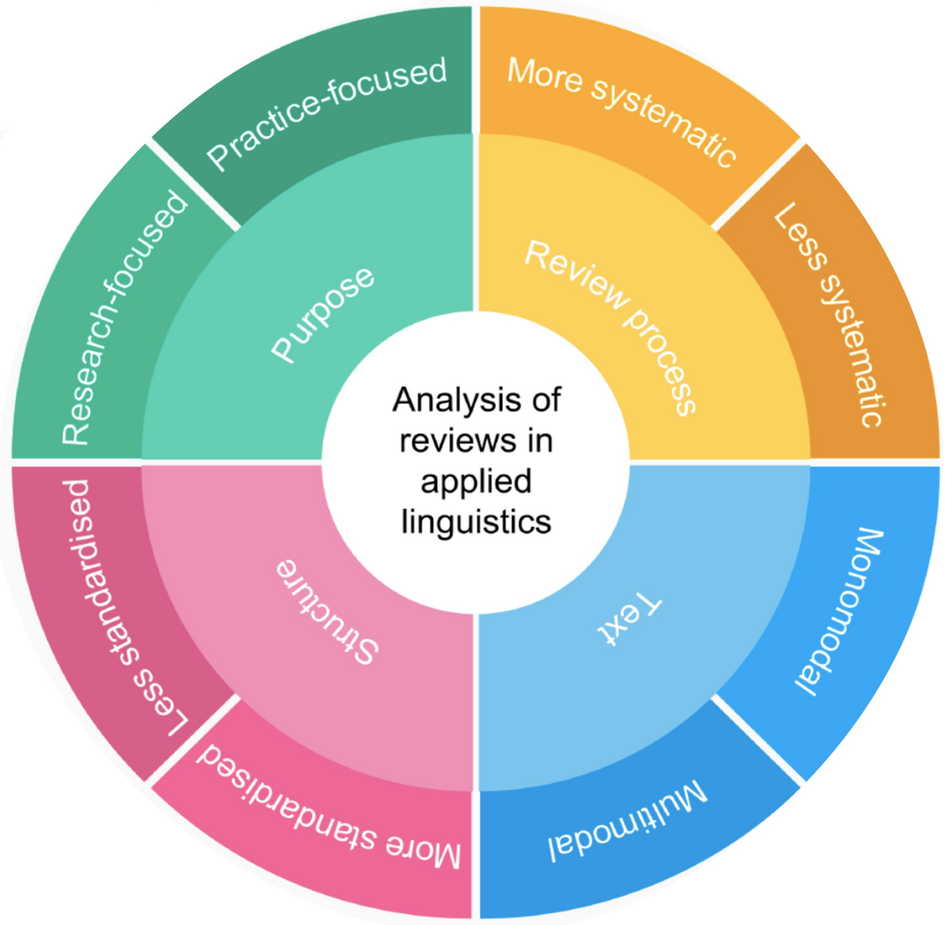
Literature review/Secondary research

Traditional (Non-systematic)

Systematic (Research synthesis)

Qualitative

Quantitative



1. critical review; 2. meta-analysis; 3. methodological synthesis; 4. mixed review; 5. narrative review;
6. qualitative research synthesis; 7. research agenda; 8. research into practice; 9. scoping review;
10. state-of-the-art review; 11. systematic literature review; 12. historical review; 13. bibliometric review

Some syntheses of DDL

Qualitative (narrative)

- 2007 Chambers (12 studies)
- 2007 Boulton (39 studies)
- 2010 Boulton (27 studies, learning outcomes)
- 2011 Yoon (12 studies, concordancing)
- 2012 Boulton (20 studies, ESP)
- **2013 Boulton & Tyne (116 studies)**
- 2017 Luo & Zhou (18 studies, writing)
- 2017 Boulton (46 studies, historical timeline)
- 2018 Chen & Flowerdew (37 studies, EAP)
- 2019 Al-Gamal & Ali (5 studies, recent)
- 2023 Sun & Park (32, collocations)

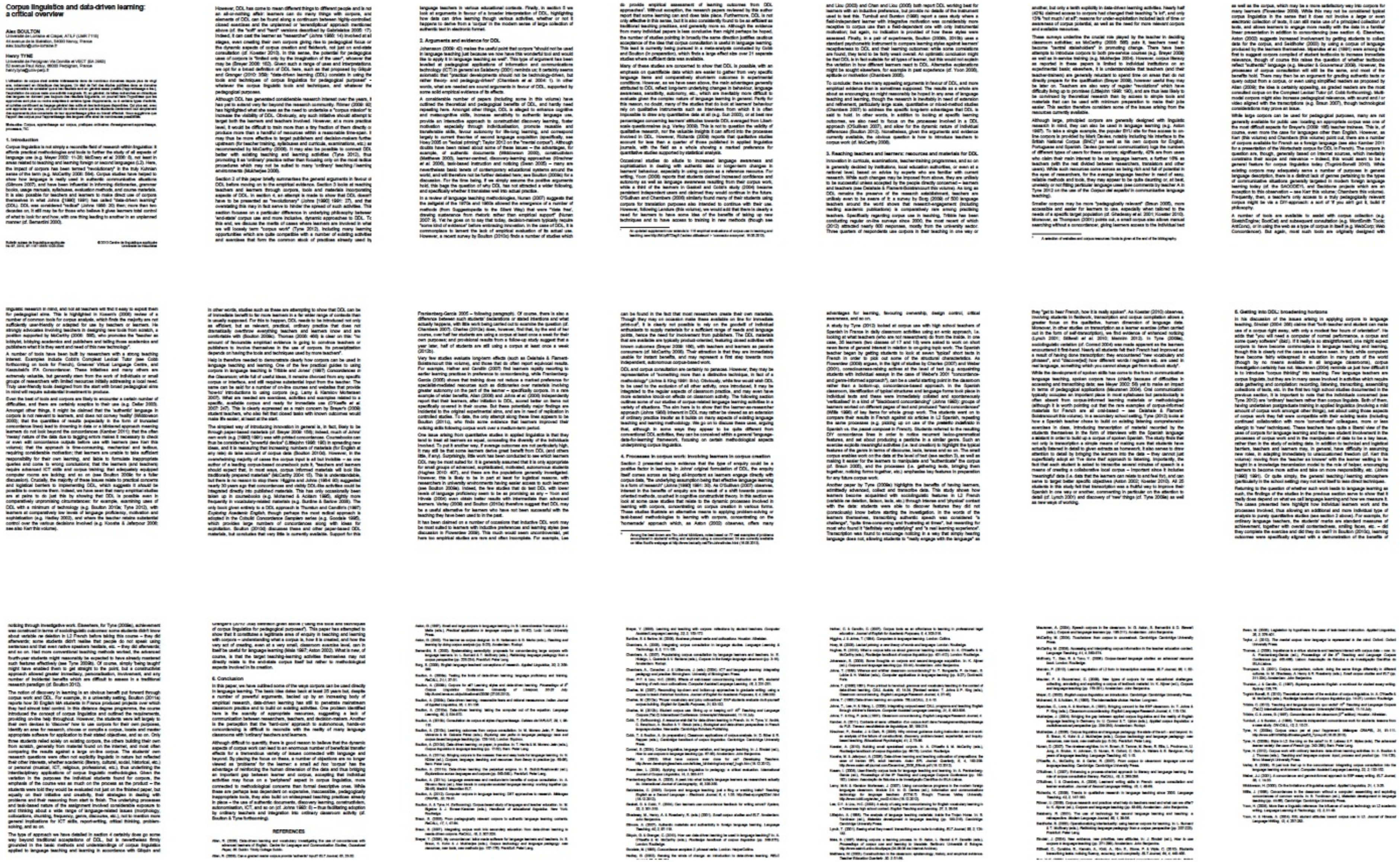
Quantitative (meta-analyses)

- 2015 Mizumoto & Chujo (14 studies, Japan)
- 2015 Cobb & Boulton (21 studies, preliminary)
- 2017 Boulton & Cobb (64 studies)
- 2019 Lee et al. (29 studies, vocab)
- 2023 Ueno & Takeuchi (144 studies)

Other (mixed)

- 2019 He & Wei (328 studies, bibliometric)
- 2021 Boulton (351 studies, coding)
- 2021 Boulton & Vyatkina (489 studies, scoping)
- 2022 Pérez-Paredes (32 studies, keywords/clusters)
- 2023 Dong et al. (126 studies, bibliometric)
- 2023 Lusta et al. (89 studies, systematic review)
- 2024 Boulton & Vyatkina (148 studies, English, JIF)

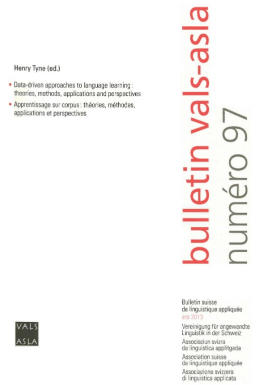
Narrative synthesis (Boulton & Tyne 2013)



Corpus linguistics and data-driven learning: a critical overview

Alex BOULTON

Henry TYNE



Henry Tyne (ed.)

- Data-driven approaches to language learning: theories, methods, applications and perspectives
- Apprentissage sur corpus: théories, méthodes, applications et perspectives

bulletin vals-asla
numéro 97

VALS ASLA

Bulletin suisse de linguistique appliquée
été 2013

Vereinigung für angewandte Linguistik in der Schweiz
Associaziun svizra da linguistica applitgada
Association suisse de linguistique appliquée
Associazione svizzera di linguistica applicata

☺ Wide-ranging, rich, in-depth

☹ Cherry-picking, subjective

Some syntheses of DDL

Qualitative (narrative)

- 2007 Chambers (12 studies)
- 2007 Boulton (39 studies)
- 2010 Boulton (27 studies, learning outcomes)
- 2011 Yoon (12 studies, concordancing)
- 2012 Boulton (20 studies, ESP)
- 2013 Boulton & Tyne (116 studies)
- 2017 Luo & Zhou (18 studies, writing)
- 2017 Boulton (46 studies, historical timeline)
- 2018 Chen & Flowerdew (37 studies, EAP)
- 2019 Al-Gamal & Ali (5 studies, recent)
- 2023 Sun & Park (32, collocations)

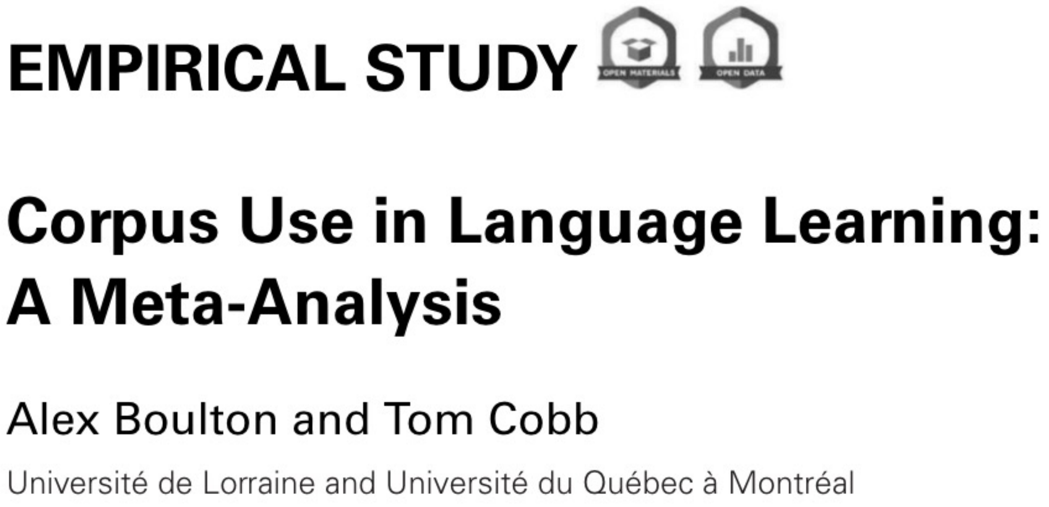
Quantitative (meta-analyses)

- 2015 Mizumoto & Chujo (14 studies, Japan)
- 2015 Cobb & Boulton (21 studies, preliminary)
- **2017 Boulton & Cobb (64 studies)**
- 2019 Lee et al. (29 studies, vocab)
- 2023 Ueno & Takeuchi (144 studies)

Other (mixed)

- 2019 He & Wei (328 studies, bibliometric)
- 2021 Boulton (351 studies, coding)
- 2021 Boulton & Vyatkina (489 studies, scoping)
- 2022 Pérez-Paredes (32 studies, keywords/clusters)
- 2023 Dong et al. (126 studies, bibliometric)
- 2023 Lusta et al. (89 studies, systematic review)
- 2024 Boulton & Vyatkina (148 studies, English, JIF)

A meta-analysis of DDL (Boulton & Cobb 2017)



EMPIRICAL STUDY

Corpus Use in Language Learning: A Meta-Analysis

Alex Boulton and Tom Cobb

Université de Lorraine and Université du Québec à Montréal

LANGUAGE LEARNING

To see:

a) *if* DDL works
b) *how well* DDL works
c) *where* DDL works (…or doesn't)

☺Quantitative: rigorous, pooled data for clear answers

☹Quantitative: less inclusive, less nuanced, subjective

Results – and SLA

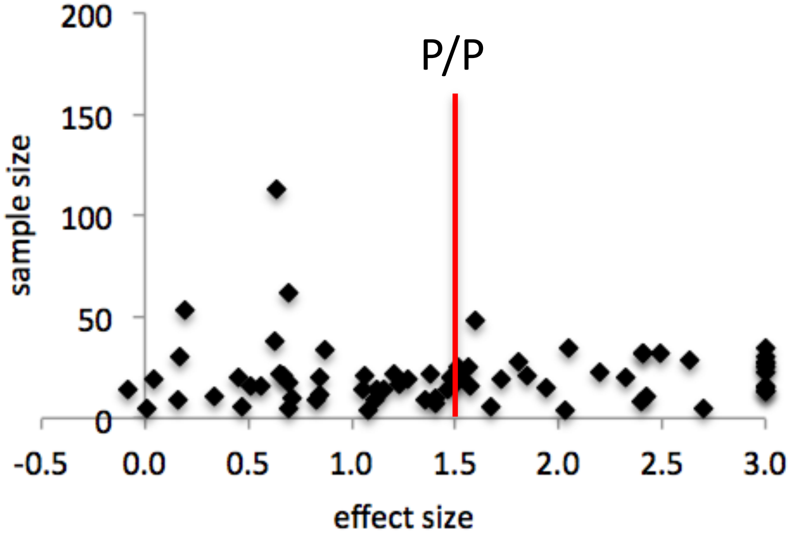
$$d = \frac{M_2 - M_1}{\sqrt{\frac{SD_1{}^2 + SD_2{}^2}{2}}}$$

Effect size	Plonsky & Oswald 2014 (C/E, *n*=67)	Plonsky & Oswald 2014 (P/P, *n*=25)	cf. SLA
large	0.9	1.4	1st quartile
medium	0.6	1.0	2nd quartile
small	0.4	0.6	3rd quartile

Boulton & Cobb 2017	0.95 (*k*=50)	1.50 (*k*=71)

☺☺☺☺☺☺☺☺

DDL large effects. DDL good.
End of story. Everyone go home.



Moderator Variables:

“DDL works pretty well in almost any context where it has been extensively tried.” (p. 386)

But…

Some syntheses of DDL

Qualitative (narrative)

- 2007 Chambers (12 studies)
- 2007 Boulton (39 studies)
- 2010 Boulton (27 studies, learning outcomes)
- 2011 Yoon (12 studies, concordancing)
- 2012 Boulton (20 studies, ESP)
- 2013 Boulton & Tyne (116 studies)
- 2017 Luo & Zhou (18 studies, writing)
- 2017 Boulton (46 studies, historical timeline)
- 2018 Chen & Flowerdew (37 studies, EAP)
- 2019 Al-Gamal & Ali (5 studies, recent)
- 2023 Sun & Park (32, collocations)

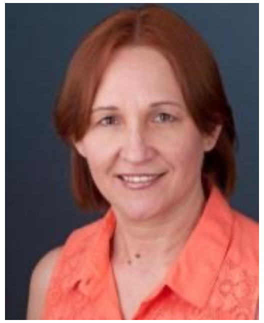
Quantitative (meta-analyses)

- 2015 Mizumoto & Chujo (14 studies, Japan)
- 2015 Cobb & Boulton (21 studies, preliminary)
- 2017 Boulton & Cobb (64 studies)
- 2019 Lee et al. (29 studies, vocab)
- 2023 Ueno & Takeuchi (144 studies)

Other (mixed)

- 2019 He & Wei (328 studies, bibliometric)
- 2021 Boulton (351 studies, coding)
- **2021 Boulton & Vyatkina (489 studies, scoping)**
- 2022 Pérez-Paredes (32 studies, keywords/clusters)
- 2023 Dong et al. (126 studies, bibliometric)
- 2023 Lusta et al. (89 studies, systematic review)
- **2024 Boulton & Vyatkina (148 studies, English, JIF)**

RA Collection



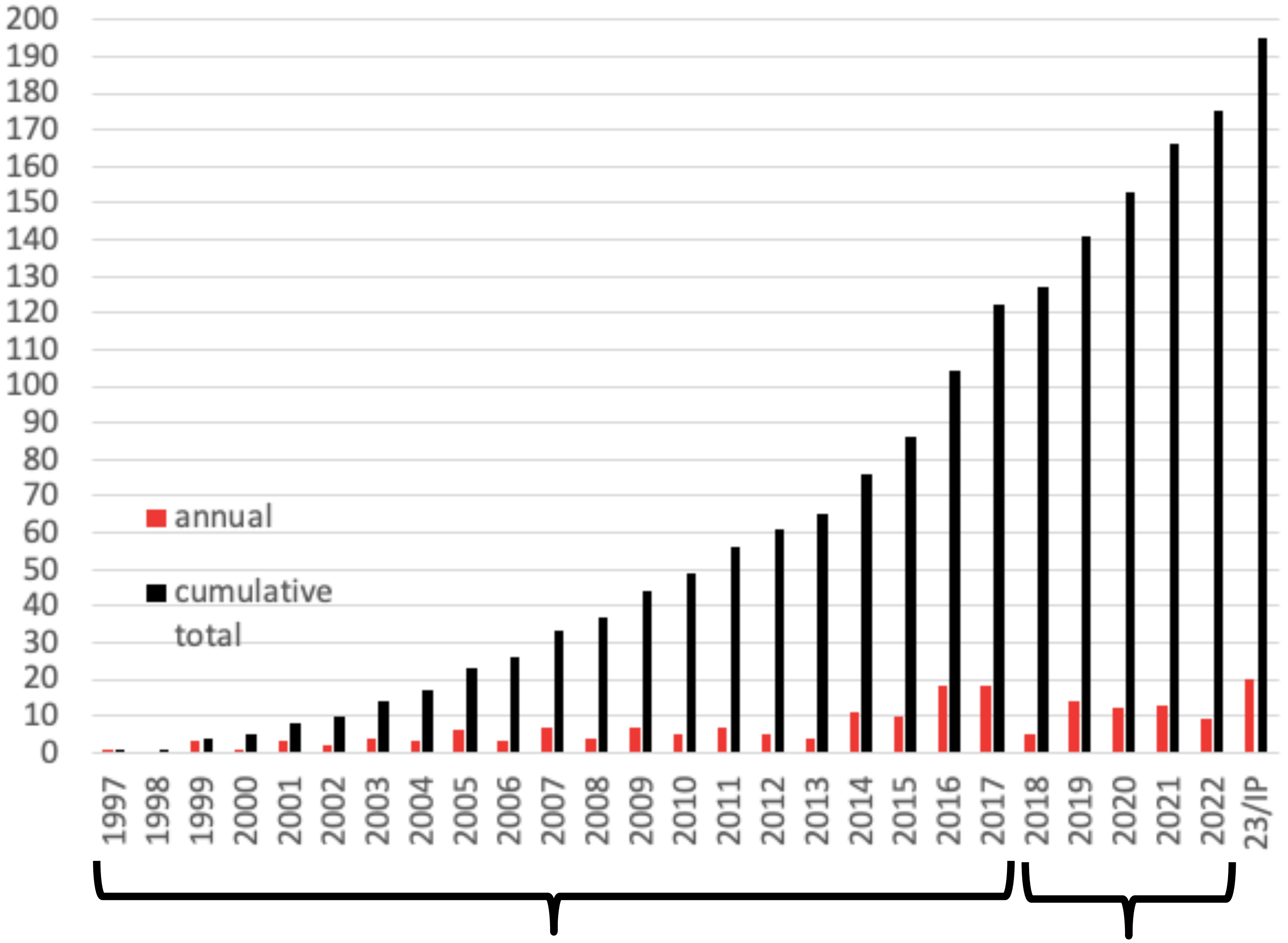
- Methodical collection of published empirical DDL studies
  (cf. Boulton & Cobb, 2017; Boulton, 2021; Boulton & Vyatkina, 2021; Boulton & Vyatkina, 2024)
- Today: up to 2022 inclusive (thanks to A. Jakob Johnson)
  DDL, empirical, in English, JCR-ranked LING+EDU journals
  ☺ ±exhaustive, but… ☹ what's NOT included
  ☺ highly visible, but… ☹ impact factor ≠ not quality! Clarivate Analytics

In the last 5 years (2018-2022):

RQ1. What trends are emerging in DDL research?
coding and analysis – manual

RQ2. How do researchers talk about DDL?
corpus analysis ('aboutness') – AntConc

RA corpus & timeline



122 RAs

53 RAs

Title	1997-2017	2018-2022	TOTAL
ReCALL	26	6	32
CALL	24	3	27
LLT	23	2	25
System	8	6	14
JEAP	3	6	9
IJAL	3	5	8
ESP	4	3	7
ELTJ	5	1	6
IJLex	4	2	6
JSLW	3	2	5
JCAL	3		3
Lawareness	3		3
BJET	1	1	2
EIT	2		2
ETS	2		2
ILE	2		2
JCHE	2		2
LTR	2		2
MLJ	1	1	2
Perspective	1	1	2
RELC Journal	2		2
Misc.	12		12
	122	**53**	**175**

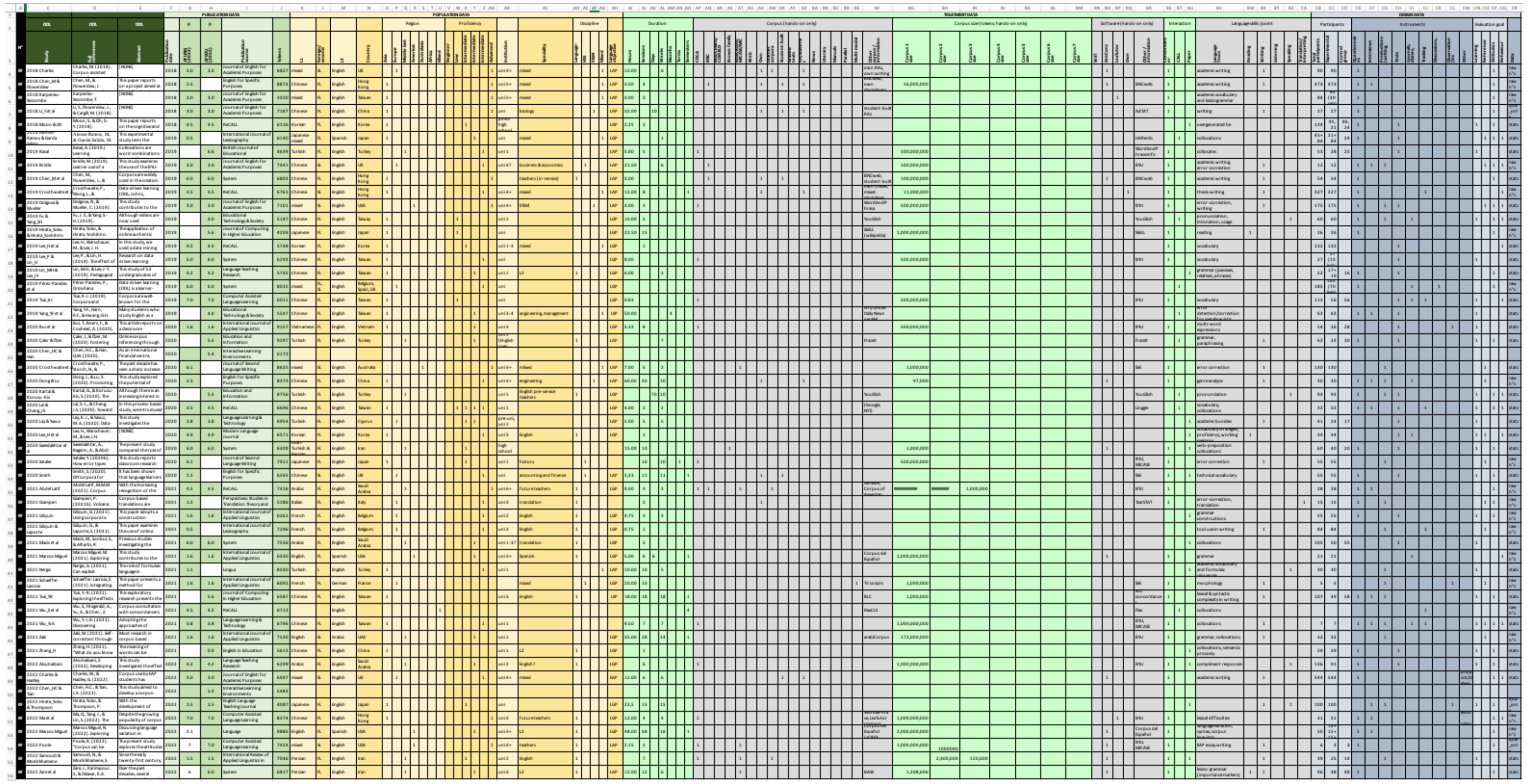
RQ1 Coding

Coding sheet

(cf. B&V 2021 in IRIS repository)

Excerpt JIF 2018-2022

IRR: decisions, decisions…



- **Publication:** ID, reference, abstract, date, JIF, source, tokens
- **Population:** L1, FL/SL, L2, country, region, proficiency, institution, speciality, discipline, LGP…
- **Treatment:** duration, corpora, size (hands-on), software, interaction, item/skills
- **Research design:** sample, instruments, objective (L/R/A/B), data (Q/Q)

RQ1 Coding: 1997-2016 ➪ 2017-2021

Size (main corpus, hands on):

<1m	31% ➪ 4%
1<99m	36% ➪ 28%
>100m	32% ➪ 68%

Variety (hands-on only) today:

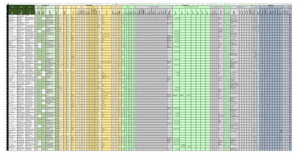
- 0 graded, news, literary, textbooks, parallel
- 1 multimodal

Skills (identifiable, multiple):

writing	56% ➪ 88%
reading	16% ➪ 12%
speaking	5% ➪ 20%
listening	2% ➪ 0%
translation	21% ➪ 4%

Language focus (identifiable, multiple):

vocabulary	24% ➪ 27%
lexicogrammar	34% ➪ 37%
grammar	16% ➪ 12%
discourse	10% ➪ 10%
correction	15% ➪ 15%



RQ2. Corpus

Corpus analysis

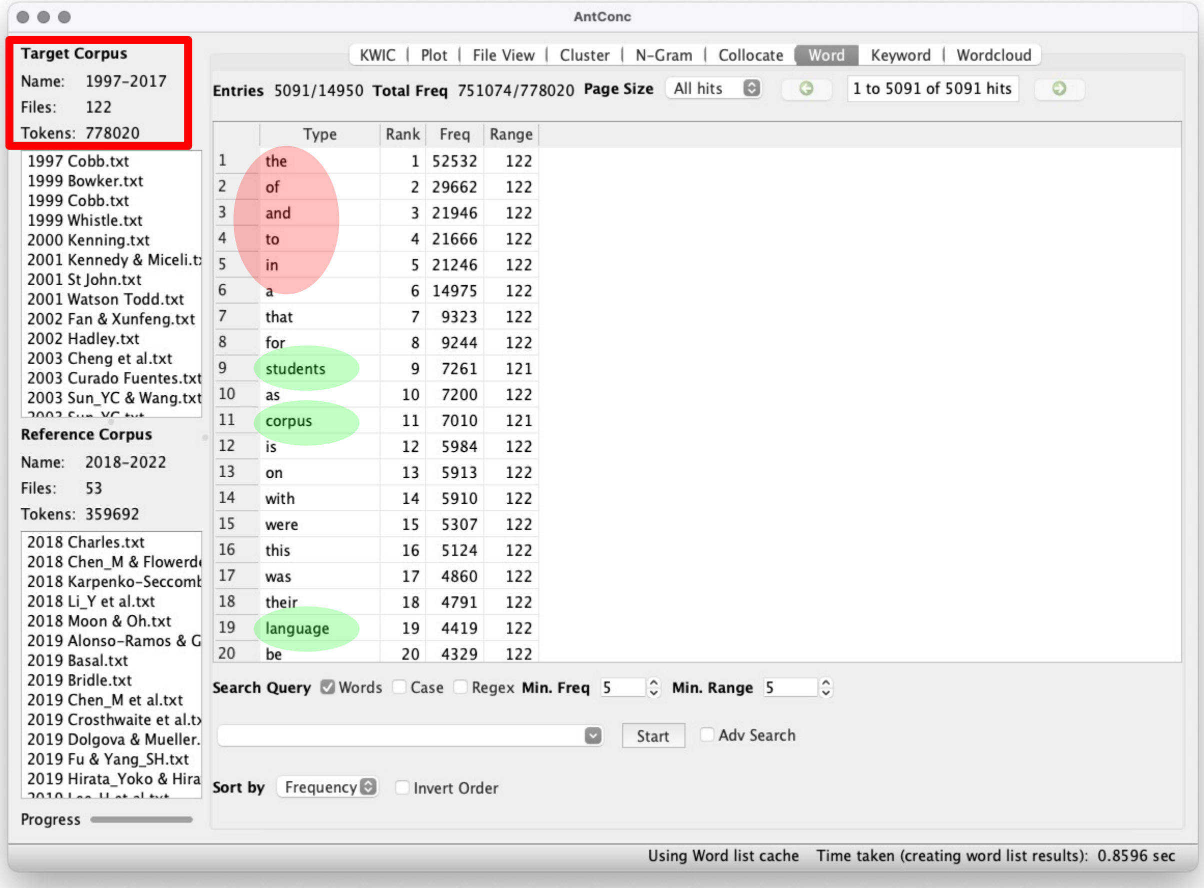
v4.2.4 + v3.5.8 (Anthony, 2023, 2019)
https://www.laurenceanthony.net/software/antconc

See also:
Jablonkai, R.R., Kim, J., & Yan, R. (in press). A corpus approach to systematic literature reviews. In K. Sadeghi (Ed.), *Routledge handbook of technological advances in researching language learning*.

- 175 RAs ➪ AntFileConverter ➪ txt (UTF8)
- main text (meta-data, headers/footers, figs/tables, extracts, foot/endnotes, references, appendices, acknowledgements, etc.)
- check! (hyphens, ligatures; X errors)

	1997-2017	2018-2022	TOTAL
papers	122	53	175
tokens	778,020	359,692	1,137,712

Corpus analysis: wordlist


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud
Start
Progress

Target Corpus
Name: 1997-2017
Files: 122
Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Reference Corpus
Name: 2018-2022
Files: 53
Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Entries 5091/14950 **Total Freq** 751074/778020 **Page Size** All hits 1 to 5091 of 5091 hits

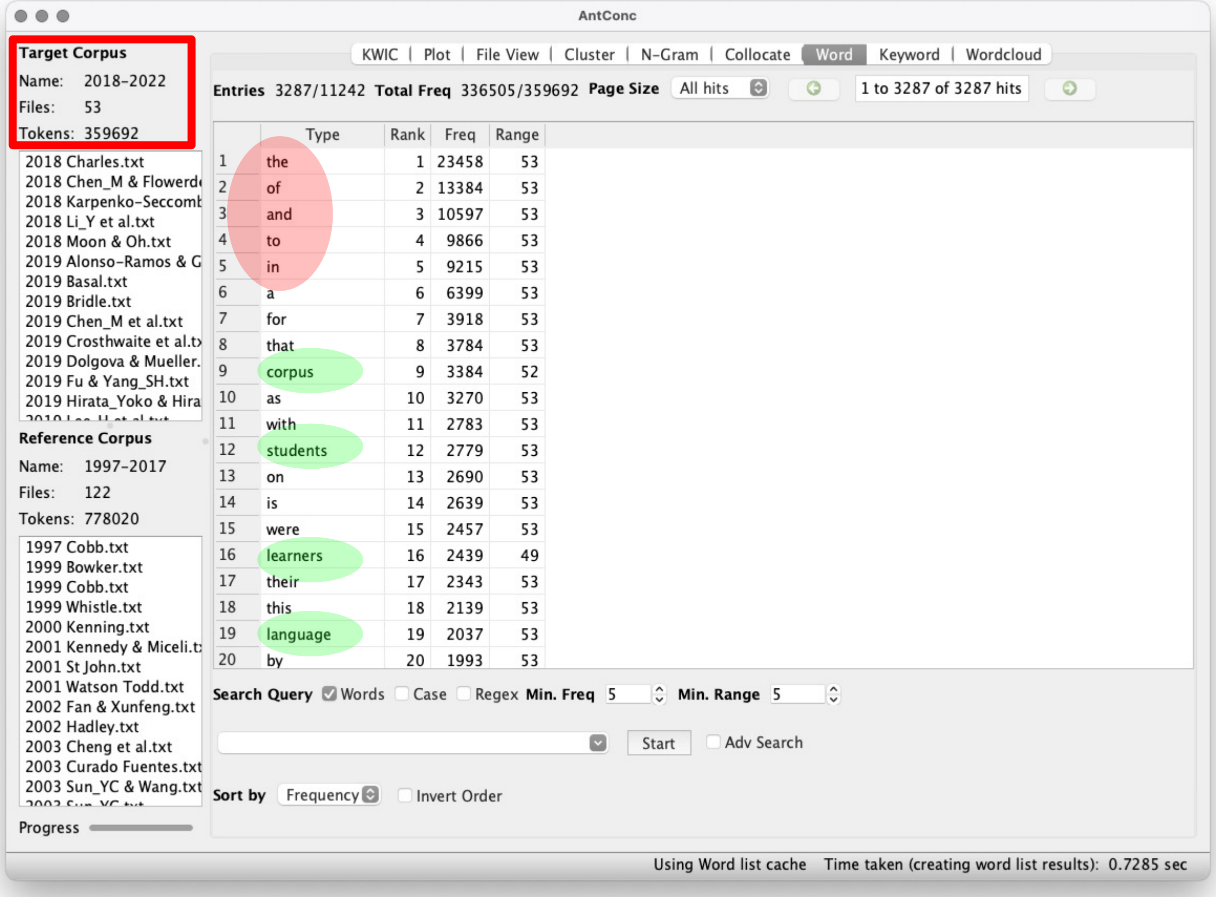
	Type	Rank	Freq	Range
1	the	1	52532	122
2	of	2	29662	122
3	and	3	21946	122
4	to	4	21666	122
5	in	5	21246	122
6	a	6	14975	122
7	that	7	9323	122
8	for	8	9244	122
9	students	9	7261	121
10	as	10	7200	122
11	corpus	11	7010	121
12	is	12	5984	122
13	on	13	5913	122
14	with	14	5910	122
15	were	15	5307	122
16	this	16	5124	122
17	was	17	4860	122
18	their	18	4791	122
19	language	19	4419	122
20	be	20	4329	122

Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

□ Adv Search

Sort by Frequency □ Invert Order

Using Word list cache Time taken (creating word list results): 0.8596 sec


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud
Start
Progress

Target Corpus
Name: 2018-2022
Files: 53
Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Reference Corpus
Name: 1997-2017
Files: 122
Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Entries 3287/11242 **Total Freq** 336505/359692 **Page Size** All hits 1 to 3287 of 3287 hits

	Type	Rank	Freq	Range
1	the	1	23458	53
2	of	2	13384	53
3	and	3	10597	53
4	to	4	9866	53
5	in	5	9215	53
6	a	6	6399	53
7	for	7	3918	53
8	that	8	3784	53
9	corpus	9	3384	52
10	as	10	3270	53
11	with	11	2783	53
12	students	12	2779	53
13	on	13	2690	53
14	is	14	2639	53
15	were	15	2457	53
16	learners	16	2439	49
17	their	17	2343	53
18	this	18	2139	53
19	language	19	2037	53
20	by	20	1993	53

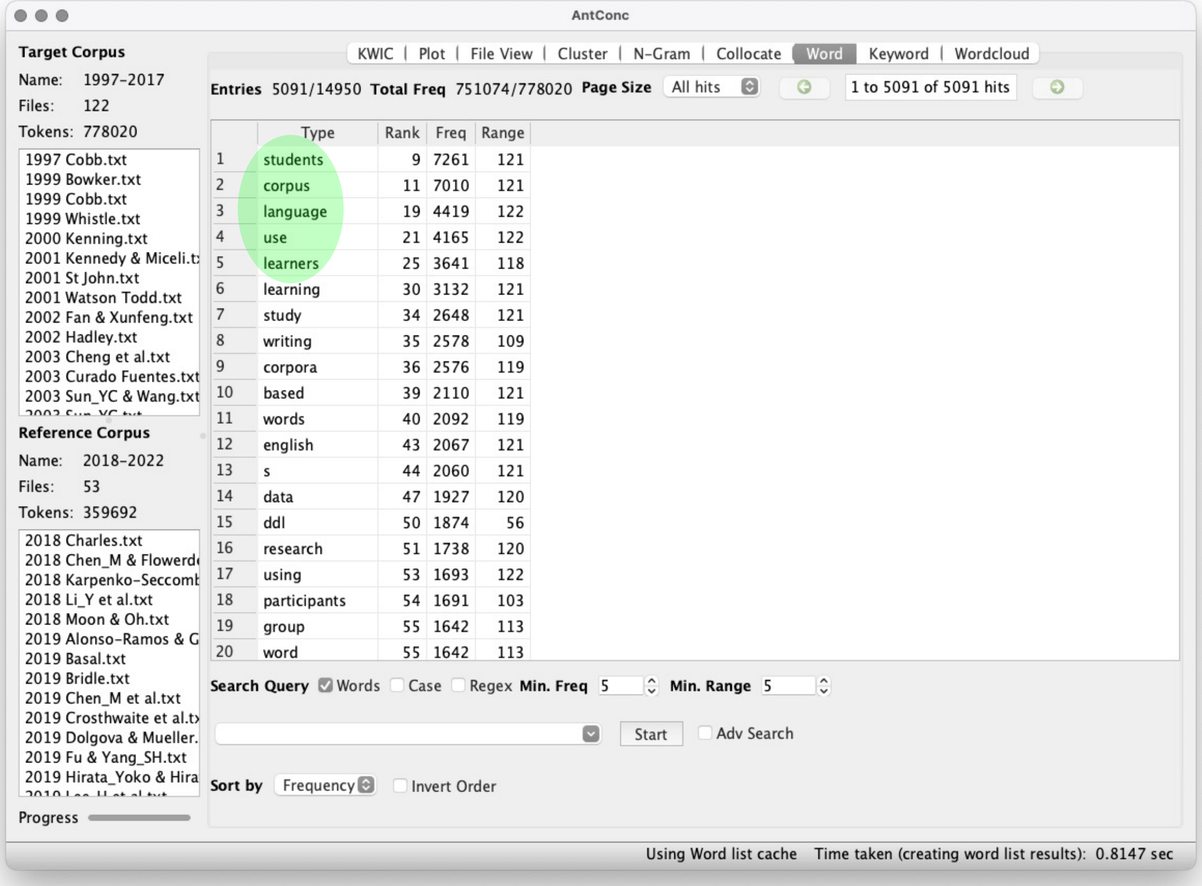
Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

□ Adv Search

Sort by Frequency □ Invert Order

Using Word list cache Time taken (creating word list results): 0.7285 sec

Corpus analysis: +stoplist


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud
Start
Progress

Target Corpus
Name: 1997-2017
Files: 122
Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Reference Corpus
Name: 2018-2022
Files: 53
Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Entries 5091/14950 **Total Freq** 751074/778020 **Page Size** All hits   1 to 5091 of 5091 hits

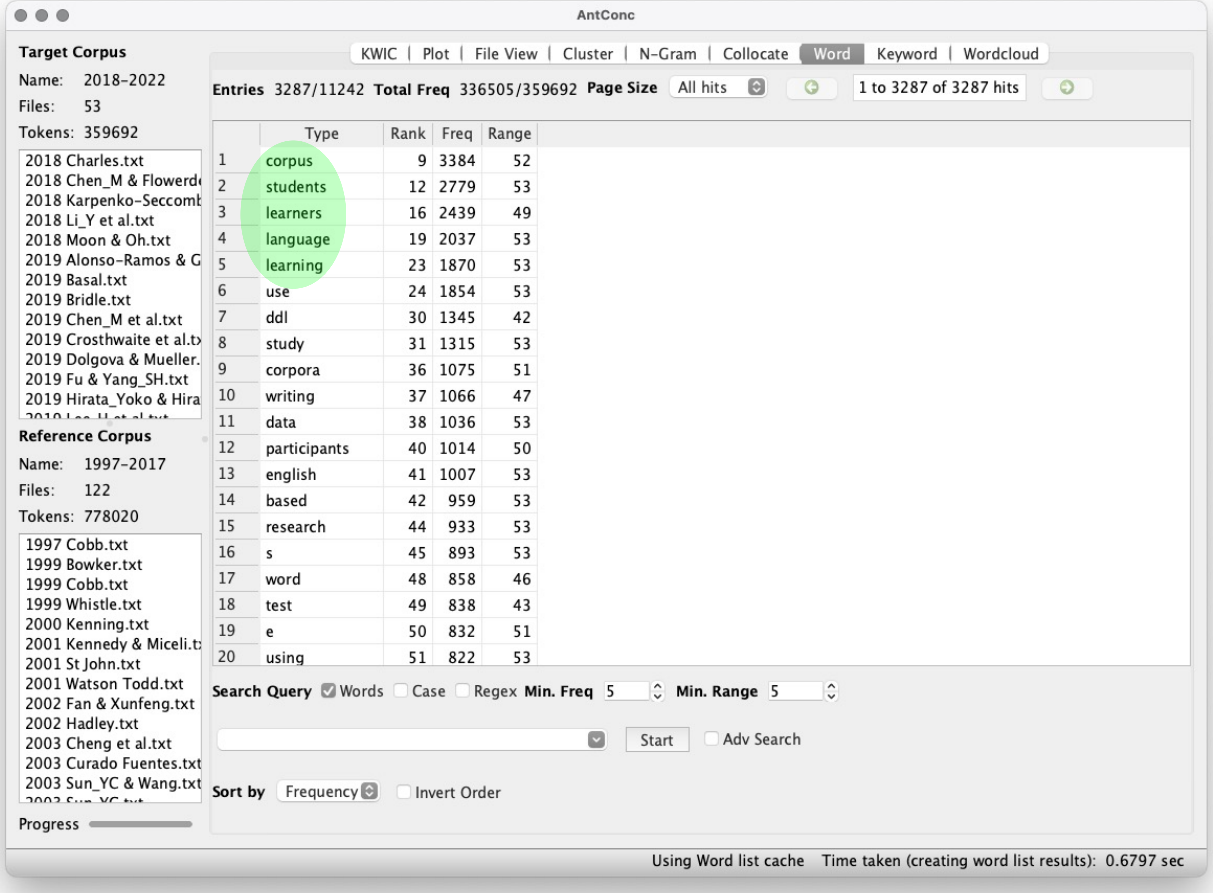
	Type	Rank	Freq	Range
1	students	9	7261	121
2	corpus	11	7010	121
3	language	19	4419	122
4	use	21	4165	122
5	learners	25	3641	118
6	learning	30	3132	121
7	study	34	2648	121
8	writing	35	2578	109
9	corpora	36	2576	119
10	based	39	2110	121
11	words	40	2092	119
12	english	43	2067	121
13	s	44	2060	121
14	data	47	1927	120
15	ddl	50	1874	56
16	research	51	1738	120
17	using	53	1693	122
18	participants	54	1691	103
19	group	55	1642	113
20	word	55	1642	113

Search Query ☑ Words ☐ Case ☐ Regex **Min. Freq** 5 **Min. Range** 5

☐ Adv Search

Sort by Frequency ☐ Invert Order

Using Word list cache   Time taken (creating word list results): 0.8147 sec


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud
Start
Progress

Target Corpus
Name: 2018-2022
Files: 53
Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Reference Corpus
Name: 1997-2017
Files: 122
Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Entries 3287/11242 **Total Freq** 336505/359692 **Page Size** All hits   1 to 3287 of 3287 hits

	Type	Rank	Freq	Range
1	corpus	9	3384	52
2	students	12	2779	53
3	learners	16	2439	49
4	language	19	2037	53
5	learning	23	1870	53
6	use	24	1854	53
7	ddl	30	1345	42
8	study	31	1315	53
9	corpora	36	1075	51
10	writing	37	1066	47
11	data	38	1036	53
12	participants	40	1014	50
13	english	41	1007	53
14	based	42	959	53
15	research	44	933	53
16	s	45	893	53
17	word	48	858	46
18	test	49	838	43
19	e	50	832	51
20	using	51	822	53

Search Query ☑ Words ☐ Case ☐ Regex **Min. Freq** 5 **Min. Range** 5

☐ Adv Search

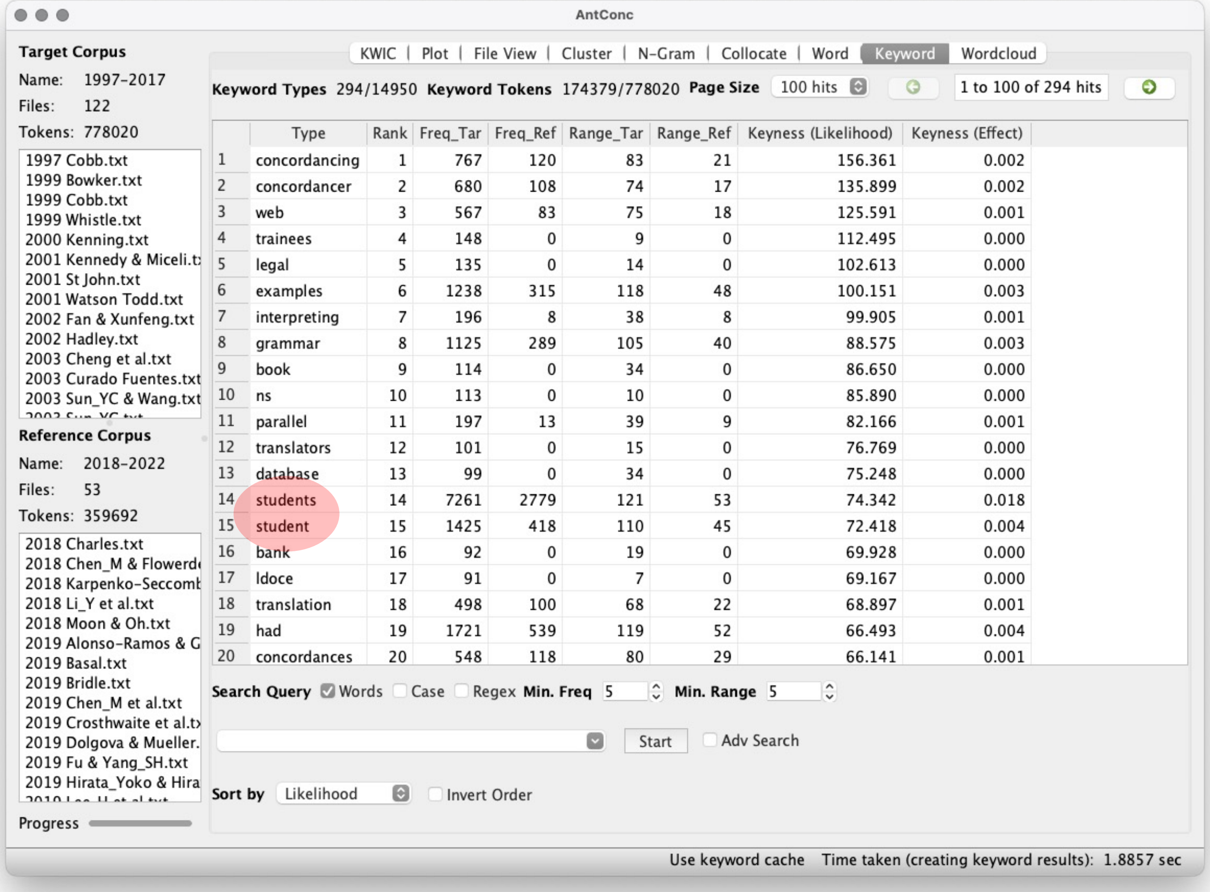
Sort by Frequency ☐ Invert Order

Using Word list cache   Time taken (creating word list results): 0.6797 sec

more revealing?

still a lot in common

Corpus analysis: keyword list



AntConc

Target Corpus

Name: 1997-2017

Files: 122

Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Reference Corpus

Name: 2018-2022

Files: 53

Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.tx
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Keyword Types 294/14950 **Keyword Tokens** 174379/778020 **Page Size** 100 hits 1 to 100 of 294 hits

	Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (Likelihood)	Keyness (Effect)
1	concordancing	1	767	120	83	21	156.361	0.002
2	concordancer	2	680	108	74	17	135.899	0.002
3	web	3	567	83	75	18	125.591	0.001
4	trainees	4	148	0	9	0	112.495	0.000
5	legal	5	135	0	14	0	102.613	0.000
6	examples	6	1238	315	118	48	100.151	0.003
7	interpreting	7	196	8	38	8	99.905	0.001
8	grammar	8	1125	289	105	40	88.575	0.003
9	book	9	114	0	34	0	86.650	0.000
10	ns	10	113	0	10	0	85.890	0.000
11	parallel	11	197	13	39	9	82.166	0.001
12	translators	12	101	0	15	0	76.769	0.000
13	database	13	99	0	34	0	75.248	0.000
14	students	14	7261	2779	121	53	74.342	0.018
15	student	15	1425	418	110	45	72.418	0.004
16	bank	16	92	0	19	0	69.928	0.000
17	ldoce	17	91	0	7	0	69.167	0.000
18	translation	18	498	100	68	22	68.897	0.001
19	had	19	1721	539	119	52	66.493	0.004
20	concordances	20	548	118	80	29	66.141	0.001

Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

Start □ Adv Search

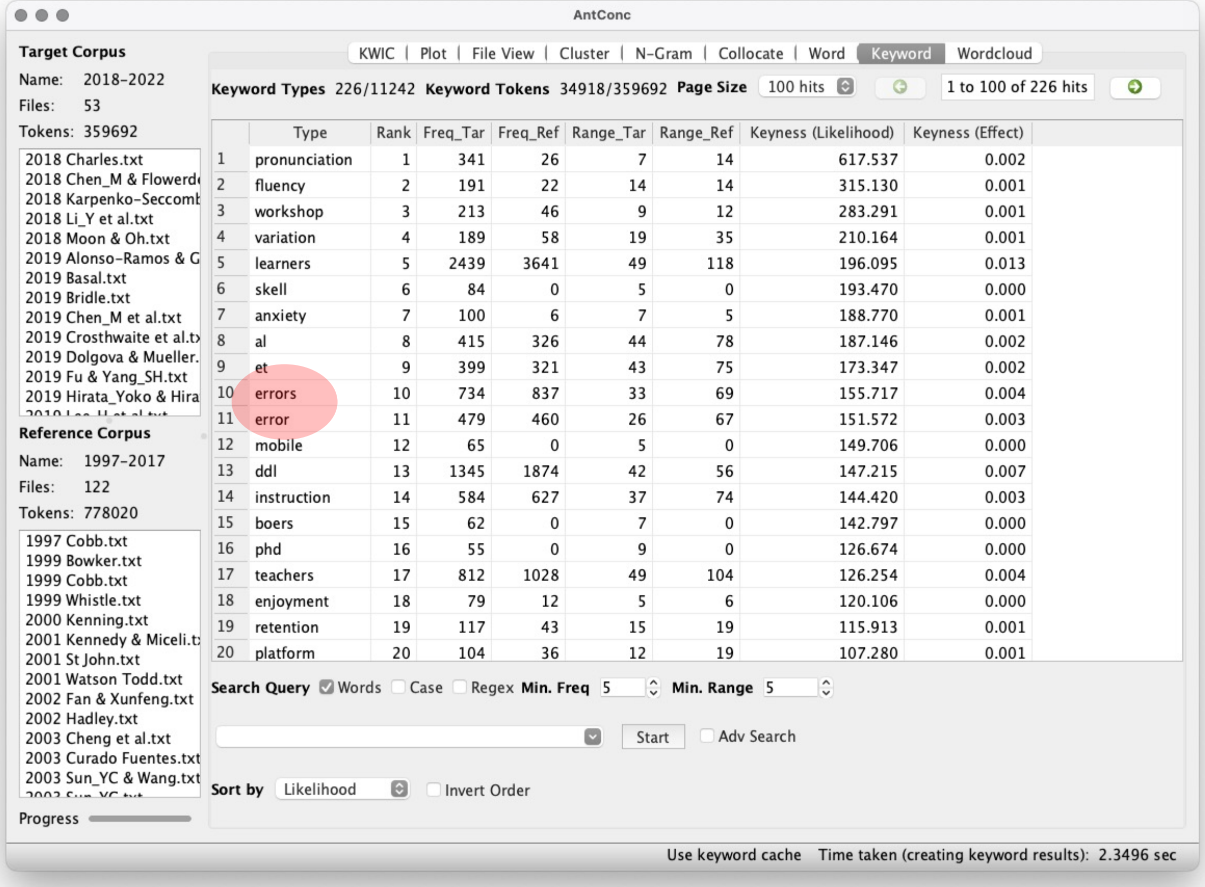
Sort by Likelihood □ Invert Order

Progress

Use keyword cache Time taken (creating keyword results): 1.8857 sec

KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud

min freq = 5, min range = 5
(both corpora)



AntConc

Target Corpus

Name: 2018-2022

Files: 53

Tokens: 359692

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.tx
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira

Reference Corpus

Name: 1997-2017

Files: 122

Tokens: 778020

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Keyword Types 226/11242 **Keyword Tokens** 34918/359692 **Page Size** 100 hits 1 to 100 of 226 hits

	Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (Likelihood)	Keyness (Effect)
1	pronunciation	1	341	26	7	14	617.537	0.002
2	fluency	2	191	22	14	14	315.130	0.001
3	workshop	3	213	46	9	12	283.291	0.001
4	variation	4	189	58	19	35	210.164	0.001
5	learners	5	2439	3641	49	118	196.095	0.013
6	skell	6	84	0	5	0	193.470	0.000
7	anxiety	7	100	6	7	5	188.770	0.001
8	al	8	415	326	44	78	187.146	0.002
9	et	9	399	321	43	75	173.347	0.002
10	errors	10	734	837	33	69	155.717	0.004
11	error	11	479	460	26	67	151.572	0.003
12	mobile	12	65	0	5	0	149.706	0.000
13	ddl	13	1345	1874	42	56	147.215	0.007
14	instruction	14	584	627	37	74	144.420	0.003
15	boers	15	62	0	7	0	142.797	0.000
16	phd	16	55	0	9	0	126.674	0.000
17	teachers	17	812	1028	49	104	126.254	0.004
18	enjoyment	18	79	12	5	6	120.106	0.000
19	retention	19	117	43	15	19	115.913	0.001
20	platform	20	104	36	12	19	107.280	0.001

Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

Start □ Adv Search

Sort by Likelihood □ Invert Order

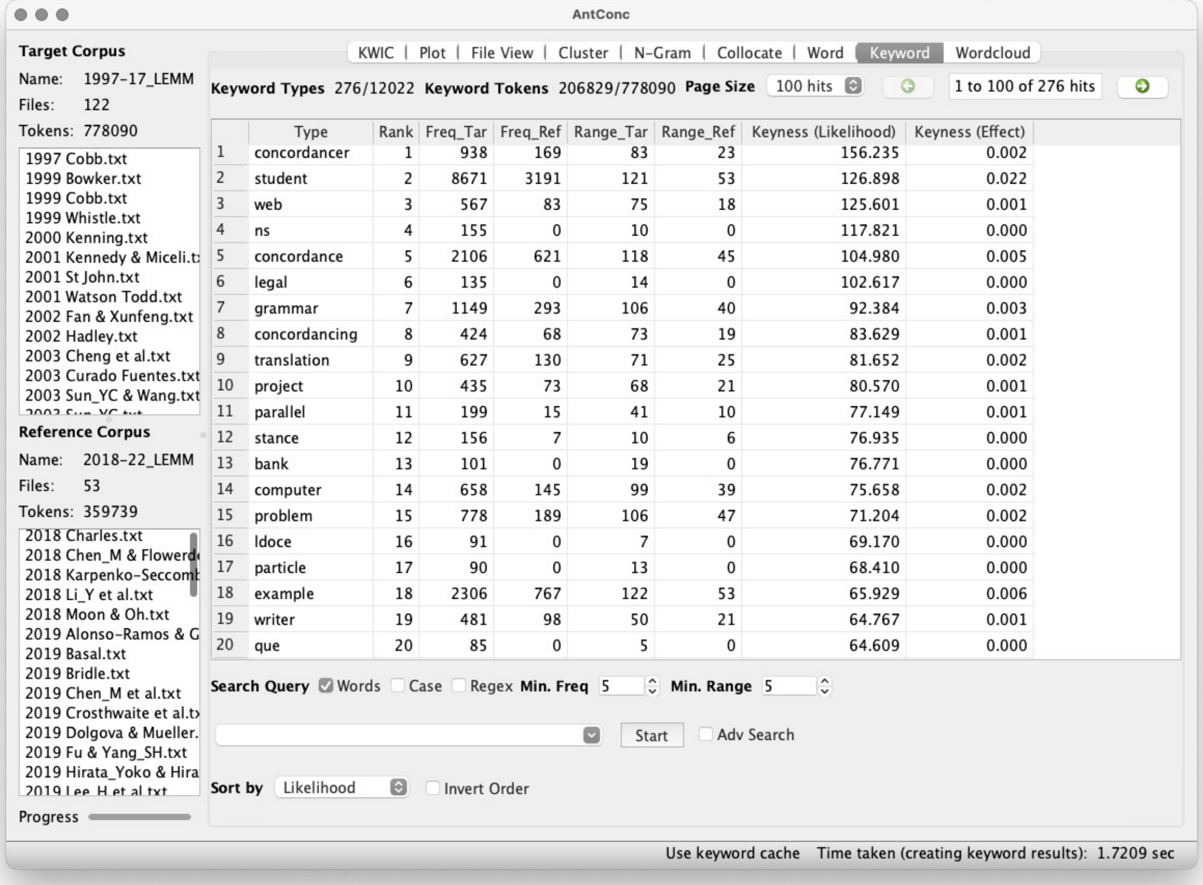
Progress

Use keyword cache Time taken (creating keyword results): 2.3496 sec

KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud

Corpus analysis: key lemmas

min freq = 5, min range = 5
(both corpora)


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud

Target Corpus
Name: 1997-17_LEMM
Files: 122
Tokens: 778090

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt

Reference Corpus
Name: 2018-22_LEMM
Files: 53
Tokens: 359739

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira
2019 Lee_H et al.txt

Keyword Types 276/12022 **Keyword Tokens** 206829/778090 **Page Size** 100 hits 1 to 100 of 276 hits

	Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (Likelihood)	Keyness (Effect)
1	concordancer	1	938	169	83	23	156.235	0.002
2	student	2	8671	3191	121	53	126.898	0.022
3	web	3	567	83	75	18	125.601	0.001
4	ns	4	155	0	10	0	117.821	0.000
5	concordance	5	2106	621	118	45	104.980	0.005
6	legal	6	135	0	14	0	102.617	0.000
7	grammar	7	1149	293	106	40	92.384	0.003
8	concordancing	8	424	68	73	19	83.629	0.001
9	translation	9	627	130	71	25	81.652	0.002
10	project	10	435	73	68	21	80.570	0.001
11	parallel	11	199	15	41	10	77.149	0.001
12	stance	12	156	7	10	6	76.935	0.000
13	bank	13	101	0	19	0	76.771	0.000
14	computer	14	658	145	99	39	75.658	0.002
15	problem	15	778	189	106	47	71.204	0.002
16	ldoce	16	91	0	7	0	69.170	0.000
17	particle	17	90	0	13	0	68.410	0.000
18	example	18	2306	767	122	53	65.929	0.006
19	writer	19	481	98	50	21	64.767	0.001
20	que	20	85	0	5	0	64.609	0.000

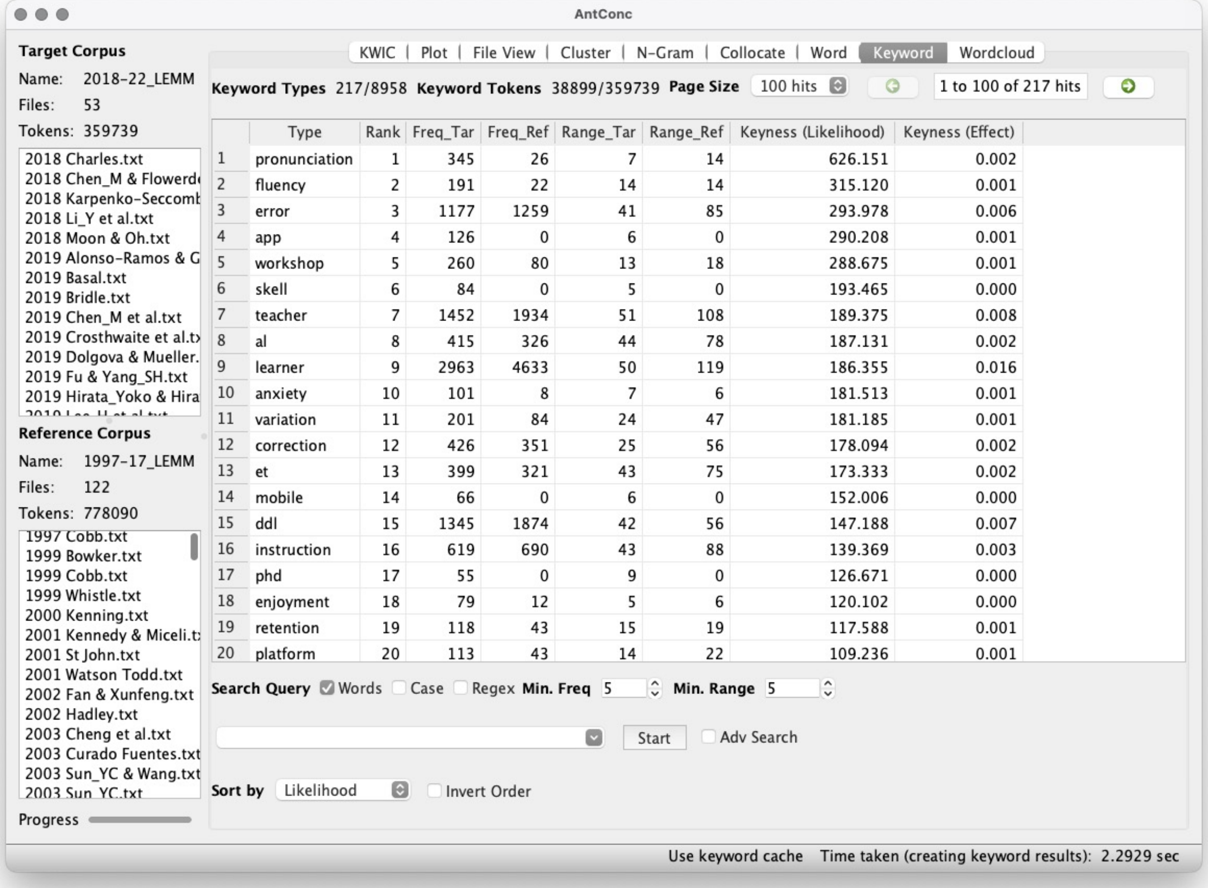
Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

Start □ Adv Search

Sort by Likelihood □ Invert Order

Progress

Use keyword cache Time taken (creating keyword results): 1.7209 sec


AntConc
KWIC | Plot | File View | Cluster | N-Gram | Collocate | Word | Keyword | Wordcloud

Target Corpus
Name: 2018-22_LEMM
Files: 53
Tokens: 359739

2018 Charles.txt
2018 Chen_M & Flowerd
2018 Karpenko-Seccomb
2018 Li_Y et al.txt
2018 Moon & Oh.txt
2019 Alonso-Ramos & G
2019 Basal.txt
2019 Bridle.txt
2019 Chen_M et al.txt
2019 Crosthwaite et al.t
2019 Dolgova & Mueller.
2019 Fu & Yang_SH.txt
2019 Hirata_Yoko & Hira
2019 Lee_H et al.txt

Reference Corpus
Name: 1997-17_LEMM
Files: 122
Tokens: 778090

1997 Cobb.txt
1999 Bowker.txt
1999 Cobb.txt
1999 Whistle.txt
2000 Kenning.txt
2001 Kennedy & Miceli.t
2001 St John.txt
2001 Watson Todd.txt
2002 Fan & Xunfeng.txt
2002 Hadley.txt
2003 Cheng et al.txt
2003 Curado Fuentes.txt
2003 Sun_YC & Wang.txt
2003 Sun_YC.txt

Keyword Types 217/8958 **Keyword Tokens** 38899/359739 **Page Size** 100 hits 1 to 100 of 217 hits

	Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (Likelihood)	Keyness (Effect)
1	pronunciation	1	345	26	7	14	626.151	0.002
2	fluency	2	191	22	14	14	315.120	0.001
3	error	3	1177	1259	41	85	293.978	0.006
4	app	4	126	0	6	0	290.208	0.001
5	workshop	5	260	80	13	18	288.675	0.001
6	skell	6	84	0	5	0	193.465	0.000
7	teacher	7	1452	1934	51	108	189.375	0.008
8	al	8	415	326	44	78	187.131	0.002
9	learner	9	2963	4633	50	119	186.355	0.016
10	anxiety	10	101	8	7	6	181.513	0.001
11	variation	11	201	84	24	47	181.185	0.001
12	correction	12	426	351	25	56	178.094	0.002
13	et	13	399	321	43	75	173.333	0.002
14	mobile	14	66	0	6	0	152.006	0.000
15	ddl	15	1345	1874	42	56	147.188	0.007
16	instruction	16	619	690	43	88	139.369	0.003
17	phd	17	55	0	9	0	126.671	0.000
18	enjoyment	18	79	12	5	6	120.102	0.000
19	retention	19	118	43	15	19	117.588	0.001
20	platform	20	113	43	14	22	109.236	0.001

Search Query ☑ Words □ Case □ Regex **Min. Freq** 5 **Min. Range** 5

Start □ Adv Search

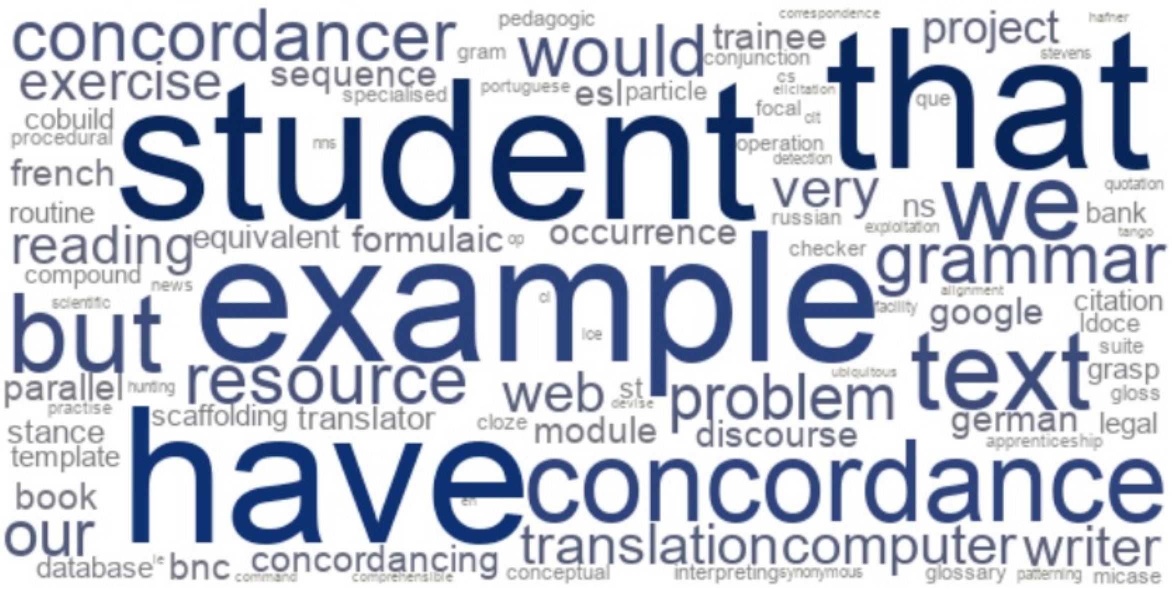
Sort by Likelihood □ Invert Order

Progress

Use keyword cache Time taken (creating keyword results): 2.2929 sec

NB how they differ
NOT what they have in common

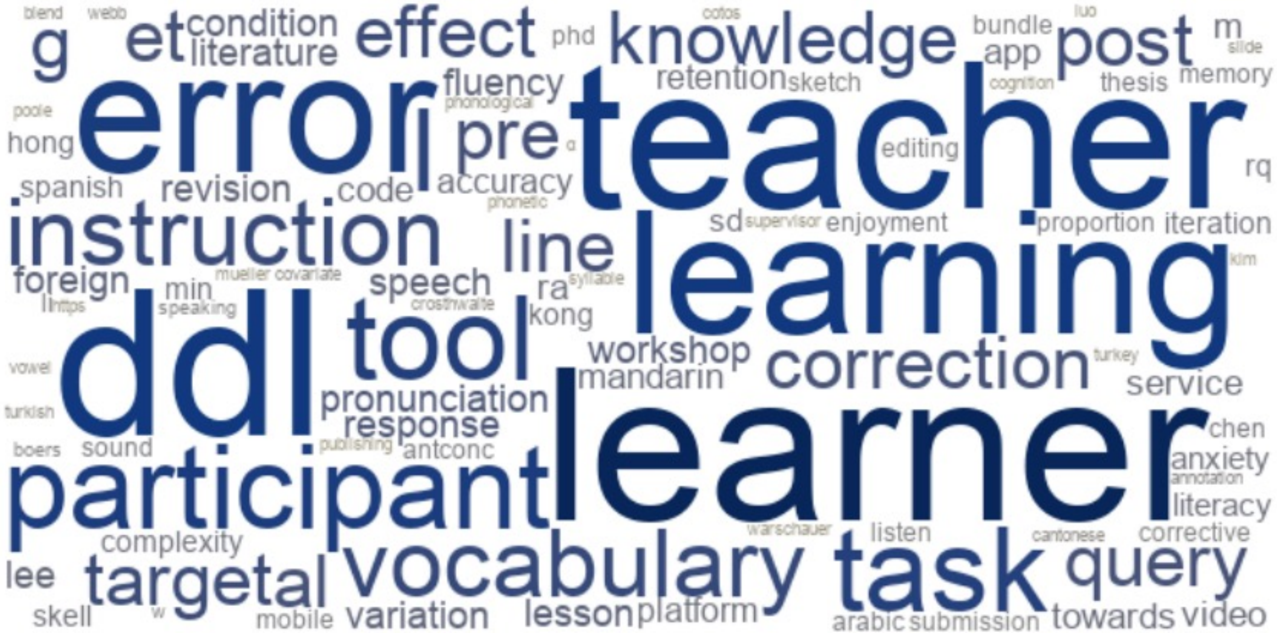
Corpus analysis: key lemmas



keylemmas 1997-2017
vs 2018-2022

min freq = 5, min range = 5
(both ways)

keylemmas 2018-2022
vs 1997-2017



Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	that
24	book	49	reading
25	esl	50	micase

themes

going down the 1997-2017 list

(principle uses)

plus key n-grams

(AntConc v3 workaround)

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

technology, tools

1997-2017	2018-2022
concordancer	ddl
concordance	query, line
concordancing	
operation	
occurrence, gram	
web, computer, google	app, mobile, platform
resource, suite, checker	tool
bank, bnc, micase,	skell
cobuild, ldoce	
book, text	

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

people involved

1997-2017	2018-2022
student, trainee	learner, phd
ns	teacher

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

(language) focus

1997-2017	2018-2022
legal, glossary, grammar particle, que, stance routine, sequence	pronunciation, fluency error, correction corrective, thesis complexity, variation vocabulary
translation, parallel interpreting writer, reading	ra
esl, french, german	mandarin, cantonese arabic, hong kong

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

activities

1997-2017	2018-2022
project, module	workshop, submission lesson, blend
problem, scaffolding	
conceptual, procedural focal	
exercise, gloss, cloze	instruction

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

some things changed
some things disappeared…

what's completely new?

themes
continuing down the 2018-2022 list

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

areas

1997-2017	2018-2022
	retention, memory
	anxiety, enjoyment

research

1997-2017	2018-2022
	al, et, boers, lee, webb crosthwaite
	iteration, rq, pre, post min, covariate, effect

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Corpus analysis: key n-grams 2018-23

Rank	3-gram (x270)
1	lee et al
2	in hong kong
3	corpus based tasks
4	of the target
5	boulton and cobb
6	the effect of
7	and learner corpora
8	and post tests
9	the pre and
10	as shown in
11	pre and post
12	in terms of
13	of corpus tools
14	the effectiveness of
15	immediate and delayed
16	the number of
17	the post test
18	and genre based
	et al p
	of the error
	the target collocations
22	the pre test
23	the concordance lines
24	the present study
25	the participants of

Rank	4-gram (x140)
1	pre and post tests
2	the pre and post
3	data driven learning ddl
4	in the pre and
5	the pre test to
6	use of corpus tools
7	in the post test
8	corpus of contemporary american
	of contemporary american english
10	as a learning tool
	between the pre and
	to use corpus tools
13	as shown in table
14	effects of ddl on
	students awareness of the
	the control and experimental
17	in english language teaching
18	as a foreign language
19	raise students awareness of
	the meanings of the
	the participants of the
	to be more effective
	in the pre test
24	in terms of the
25	engine for language learning [+10]

Rank	5-gram (x35)
1	the pre and post tests
2	the use of corpus tools
3	corpus of contemporary american english
4	in the pre and post
5	between the pre and post
	the control and experimental groups
7	sketch engine for language learning
8	the corpus of contemporary american
9	findings of the present study
	the long term effects of
	the pre test and post
12	the use of the corpus
13	on the basis of the
14	in the pre test and
	it is worth mentioning that
	of english for academic purposes
	the findings of the present
	to be more effective than
19	english as a foreign language
20	english for international communication toeic
	of language learning and teaching
	participants were randomly divided into
	the present study aims to
	the test of english for
25	data driven learning ddl johns

Corpus analysis: 50 key lemmas

1997-2017

1	concordancer	26	resource
2	student	27	operation
3	web	28	checker
4	ns	29	sequence
5	concordance	30	exercise
6	legal	31	french
7	grammar	31	bnc
8	concordancing	31	text
9	translation	34	conceptual
10	project	35	trainee
11	parallel	36	occurrence
12	stance	37	glossary
13	bank	38	german
14	computer	39	gloss
15	problem	39	scaffolding
16	ldoce	41	focal
16	particle	42	grasp
18	example	43	module
19	writer	44	procedural
20	que	45	gram
21	google	46	suite
22	interpreting	47	cloze
23	routine	48	reading
24	book	49	micase
25	esl	50	cobuild

concordancing

1997-2017		2018-2022
757	frequency	120
0.97	per thousand words	0.33
83/122 (68.0%)	range	21/53 (39.6%)

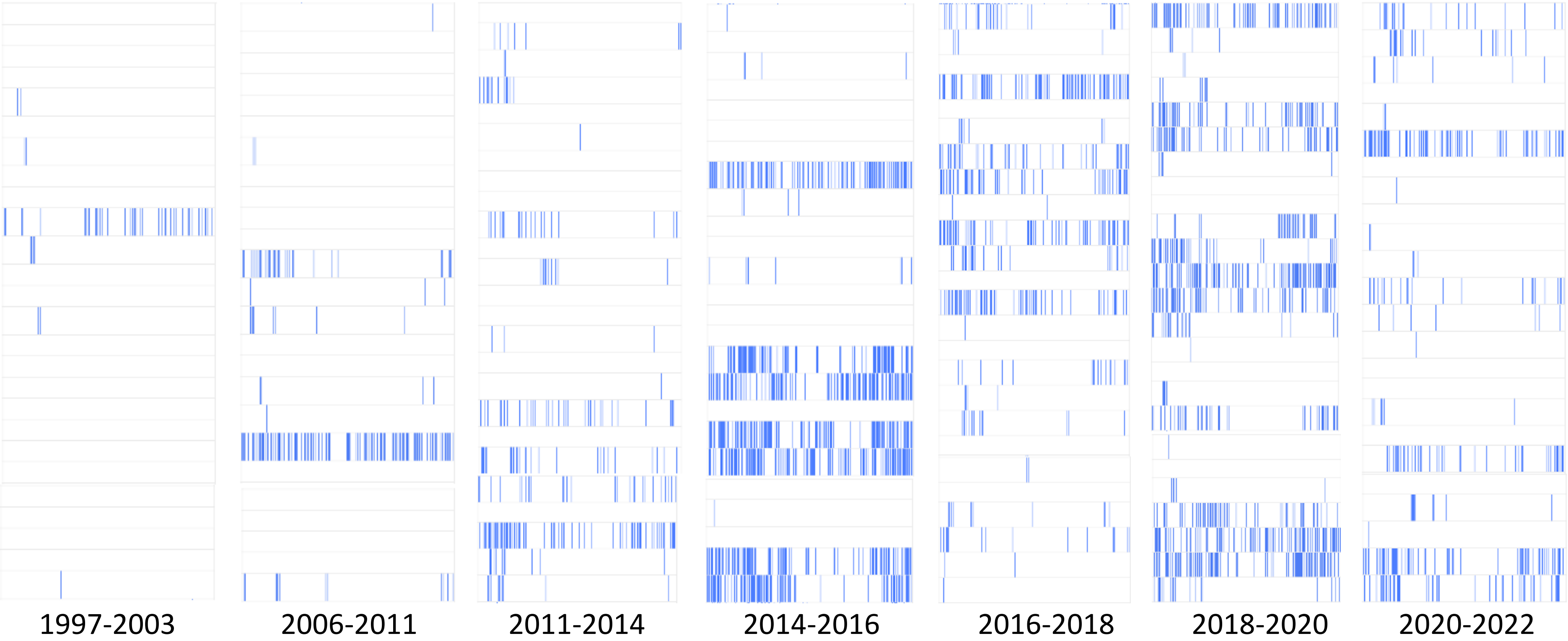
DDL

1997-2017		2018-2022
1874	frequency	1345
2.41	per thousand words	3.74
56/122 (45.9%)	range	42/53 (79.2%)

2018-2022

1	pronunciation	26	rq
2	fluency	27	lesson
3	error	28	lee
4	app	29	vocabulary
5	workshop	30	arabic
6	skell	31	post
7	teacher	31	min
8	al	31	cantonese
9	learner	34	webb
10	anxiety	35	foreign
11	variation	36	hong
12	correction	37	kong
13	et	38	covariate
14	mobile	39	ra
15	ddl	39	pre
16	instruction	41	line
17	phd	42	learning
18	enjoyment	43	query
19	retention	44	crosthwaite
20	platform	45	complexity
21	submission	46	corrective
22	mandarin	47	blend
23	thesis	48	effect
24	iteration	49	memory
25	boers	50	tool

Other corpus tools (DDL)

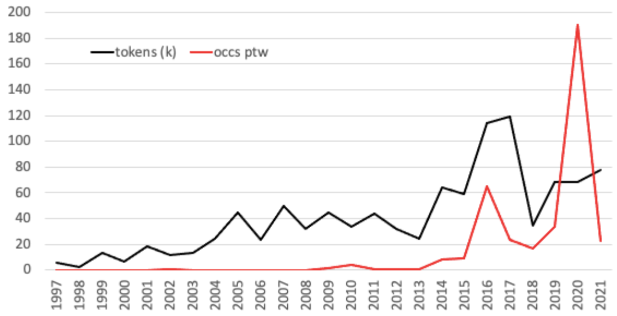


DDL over time
(inc 0 mentions)

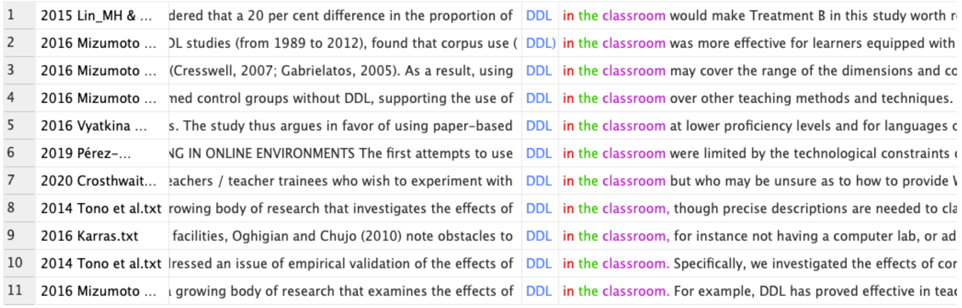
Top DDL RAs ptw

	Date RAFreq	ptw
1	2016 Mizumoto & Chujo	29.11
2	2015 Lin & Lee	27.48
3	2016 Mizumoto et al	22.46
4	2019 Lin & Lee	21.45
5	2016 Lin	19.94
6	2020 Saeedakhtar et al	17.44
7	2020 Lee et al	16.13
8	2016 Vyatkina (a)	15.61
9	2016 Vyatkina (b)	15.35
10	2018 Moon & Oh	14.72
11	2014 Smart	14.62
12	2016 Karras	14.61
13	2021 Gilquin	13.58
14	2010 Boulton	13.41
15	2017 Ackerley	12.35
16	2022 Samoudi & Modir.	11.50
17	2017 Hadley & Charles	10.61
18	2019 Crosthwaite et al	10.06

DDL: ptw



DDL: context



1	2015 Lin_MH & ...	dered that a 20 per cent difference in the proportion of	DDL	in the classroom would make Treatment B in this study worth r
2	2016 Mizumoto ...	DL studies (from 1989 to 2012), found that corpus use (	DDL)	in the classroom was more effective for learners equipped with
3	2016 Mizumoto ...	(Cresswell, 2007; Gabrielatos, 2005). As a result, using	DDL	in the classroom may cover the range of the dimensions and co
4	2016 Mizumoto ...	ned control groups without DDL, supporting the use of	DDL	in the classroom over other teaching methods and techniques.
5	2016 Vyatkina ...	s. The study thus argues in favor of using paper-based	DDL	in the classroom at lower proficiency levels and for languages c
6	2019 Pérez-...	NG IN ONLINE ENVIRONMENTS The first attempts to use	DDL	in the classroom were limited by the technological constraints (
7	2020 Crosthwait...	eachers / teacher trainees who wish to experiment with	DDL	in the classroom but who may be unsure as to how to provide \
8	2014 Tono et al.txt	rowing body of research that investigates the effects of	DDL	in the classroom, though precise descriptions are needed to cla
9	2016 Karras.txt	facilities, Oghigian and Chujo (2010) note obstacles to	DDL	in the classroom, for instance not having a computer lab, or ad
10	2014 Tono et al.txt	lressed an issue of empirical validation of the effects of	DDL	in the classroom. Specifically, we investigated the effects of cor
11	2016 Mizumoto ...	ı growing body of research that examines the effects of	DDL	in the classroom. For example, DDL has proved effective in teac

Uncertain terms: a Johnsian exegesis

(cf. Boulton, 2011)

Johns 1986. *Micro-Concord: A language learner's research tool.*

- "concordancing"

Johns 1988. *Whence and whither classroom concordancing?*

Johns 1990. *From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning*.

Johns 1991. *Should you be persuaded: Two samples of data-driven learning*.

Johns & King (eds.) 1991. *Classroom Concordancing*.

- "an application of computers to language-learning that has come to be known as 'classroom concordancing' or 'data-driven learning' (DDL)" (p.iii)

Johns 1993. *Data-driven learning: An update.*

- "The earlier term Classroom Concordancing described the technique; the new term Data-Driven Learning was coined to emphasise the methodology." (p.4)

Johns 2002. *Data-driven learning: The perpetual challenge*.

- "an approach… that I have, for want of a better term, named data-driven learning." (p.107)

Johns et al. 2008. *Integrating corpus-based CALL programs in teaching English through children's literature*.

- "corpus-based language learning" (p.495)

External comparison

175 comparison texts: ±ISLA ➪ same journal, same year (same issues); min range = 5

Rank	DDL keywords (x759)		
1	corpus	26	chambers
2	corpora	27	concordancers
3	ddl	28	patterns
4	concordance	29	deductive
5	collocations	30	bnc
6	collocation	31	word
7	concordancing	32	driven
8	concordancer	33	google
9	search	34	materials
10	concordances	35	data
11	consultation	36	linguistics
12	boulton	37	formulaic
13	use	38	based
14	searches	39	coca
15	query	40	noun
16	examples	41	lexico
17	johns	42	errors
18	lines	43	reference
19	cobb	44	phrases
20	tools	45	collocational
21	inductive	46	approach
22	hands	47	kennedy
23	yoon	48	verb
24	collocates	49	exercises
25	queries	50	preposition

Rank	Non-DDL keywords (x1048)		
1	captions	26	planning
2	interaction	27	spanish
3	social	28	chat
4	communication	29	multimodal
5	technology	30	vowel
6	face	31	cmc
7	feedback	32	wiki
8	captioning	33	cultural
9	collaborative	34	video
10	self	35	exchange
11	comprehension	36	cf
12	peer	37	digital
13	negotiation	38	voice
14	l	39	graph
15	listening	40	blended
16	mall	41	call
17	messages	42	strategies
18	environment	43	game
19	mail	44	facebook
20	mobile	45	emotions
21	scmc		synchronous
22	practices	47	global
23	clil	48	virtual
	blog	49	reading
25	captions	50	semiotic

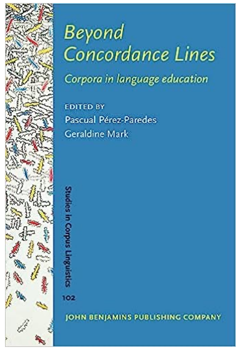
Discussion

Different syntheses (NS, MA, MM, corpus): complementary, triangulation

➪ essential to know your field! Automated, statistics, but…

1. Listen to past recommendations: better research practices, greater rigor in reporting (e.g. duration, proficiency, activities, materials)

2. More diversity, originality
   ‘corpus’ types, tools & interfaces
   AI/ChatGPT?



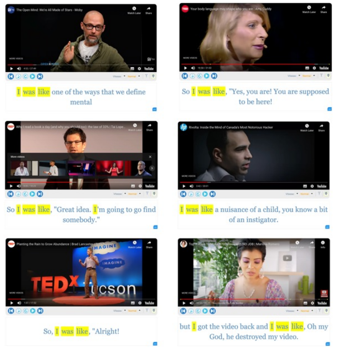
Beyond Concordance Lines

Corpora in language education

EDITED BY
Pascual Pérez-Paredes
Geraldine Mark

Studies in Corpus Linguistics
102

JOHN BENJAMINS PUBLISHING COMPANY


The Open Mind: We're All Made of Stars - Moby
Your body language may shape who you are | Amy Cuddy
Why I read a book a day (and why you should too): the law of 33% | Tai Lope...
Rivolta: Inside the Mind of Canada's Most Notorious Hacker
Planting the Rain to Grow Abundance | Brad Lancaster
TEDx Tucson

I was like one of the ways that we define mental

So I was like, "Yes, you are! You are supposed to be here!

So I was like, "Great idea. I'm going to go find somebody."

I was like a nuisance of a child, you know a bit of an instigator.

So, I was like, "Alright!

but I got the video back and I was like, Oh my God, he destroyed my video.

3. Research on the underpinnings of DDL (processes), e.g. DDL promotes autonomy, noticing, induction, language awareness …‘better learners’?

http://micase.elicorpora.info/

your program there's a form that you can mail in. um **thank you** and have a wonderful evening. APPLAUSE {END

s i learned to analyze scientific research articles later. **thank you** for not making me dread that. you have the ability to

thank you have a nice weekend UNINTELLIGIBLE CONVERSATION

uh so let's give the tape recorder a break too, and so **thank you** very much and i'll see you on Thursday. micase-related

PAUSE duration well thank you. **thank you**. {END OF TRANSCRIPT}

of getting the slides please? okay. uh there we go um, **thank you**. now look at his, another image of augustus here, um this

down and if you could just give them over to nikolas. **thank you**. UNINTELLIGIBLE SPEECH

uh i think we can just actually, stop slides, yeah **thanks** (we can get) a little more light here. um, and the scale of this

ion, right through those doors to the right, afterwards, **thanks**, for coming everyone APPLAUSE {END OF TRANSCRIP

ions before we wrap up...? okay, that concludes it then **thanks**. {END OF TRANSCRIPT}

ery nice. any questions? okay **thanks**. okay. all righty um what I

, highly intensive coffee plantations. SLIDE CHANGE **thanks**. so, given this context, then uh obviously one of the things

's a bunch of extras here. oh **thanks**. PAUSE WHILE LOWERING SCREEN so, again this is one of

m, well thank you very much. i think we're done, and **thanks** for, allowing this to be videotaped, this project thanks you.

wrry about things that we haven't discussed at all. so, **any questions** (coming up?) everyone's is th

cover on aquifer evaluation tests but i i are there **any questions**? is everyone i, you can't learn all the all the details

ore we get going with the selection sort again are there **any questions** about anything...? okay. well what i'd like to do first...

o do the exchange in the other array. kay well are there **any questions** about this? PAUSE duration :05 kay well let's start

i'm gonna assign uh practice problems for homework. **any questions** before we wrap up...? okay, that concludes it then

stions you, make sure that if you have any concerns, **any questions** email me. and what would be better is if you can

yes i'll entertain **any questions** i'm dying to ask you a questi