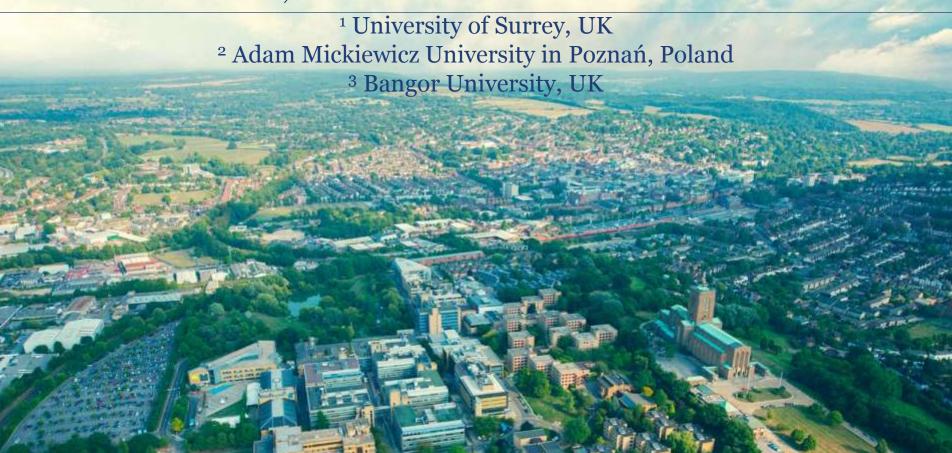


New directions in DDL for EAP writers

Ana Frankenberg-Garcia¹, Robert Lew², Geraint Paul Rees¹, Jonathan C. Roberts³ & Nirwan Sharma³



From DL to DDL



Discovery learning (DL)

- Constructivist approach (Piaget 1936, Vygotsky 1978)
 - Discovering and understanding
 - instead of rote learning

Enter corpora

sult can be found across the three corpora and over all data sizes. As with the aut based on the analysis of large text corpora often throw up interesting and unexpec rom statistics obtained from large corpora of language. The match of proximal cu view of the relationships between corpora, pedagogical lexicography and language ation of spatial phenomena in text corpora . This article is concerned with ogical dictionary. Finally, learner corpora can be used to identify deficient use of ne same reference corpus is used, corpora of different sizes can be meaningfully uses that can be made of language corpora for teaching. Practising teachers will n the urology team as aspiration of corpora is needed. Rarely, surgery may be req d from learner and native speaker corpora. The processing burden on native speal

- Discovery learning through concentrated language data
 - Data driven-learning (DDL) (Johns 1991)



Two classic approaches to DDL



The hard approach vs. the soft approach (Gabrielatos 2005)



The hands-on approach vs. the hands-off approach (Boulton 2010)



Hands-on







1) 1)Read the concordance lines below and try to guess what the missing word is.
at to try and sort of figure it all out in my and she started trying to teach me not quite clear hadn't got that clear in my Three o'clock and seven o'clock how can I Now this may just be me going off my sitting in front of Newsnight he became ultimately the erm [tc text=pause] of the careers service in London University. at one stage and took on MX who is now I think of political science erm to do [tc text=pause] cost yeah. [M01] Yes. Yeah. [F01] The deputy What kind [M01] Mm. [F01] of role does she it and FX had this terrific weight behind her and she I think she did twenty er sit ups with I was a child I could not get this out of my I cried at night with it going over and over up smashing the salad-cream bottle er over her and [tc text=pause] I didn't want to hurt her them down my neck. I was smashed out of my When I did eventually get back to the prison
a lot of people er never it never entered theirto think about these things. [M01] Really? [F01] Mm. [F02] able to perhaps with a clearer approach [F01] Mm. [F02] the more sensitive to say erm [tc text=pause] It's gone out of my [F01] Right. Erm [F02] [ZF1] I th [ZF0] I [M01] MX [M02] there was MX who was then of the Woodlands Unit [M01] Oh was he. [M02] Yeah. [M01] that they've g Well the deputy [M02] [tc text=coughs] [M01] of curriculum erm
<pre>[M01] Erm [tc text=pause] not off the top of my No [F01] Okay. Well just to conclude may I [tc text=pause] I'm not putting ideas in your but smoking seems to be associated with these you've got to have eyes in the back of your I mean [F01] Mm [F02] having kids yourself you s [ZF1] the [ZF0] the head is he [M01] He's the of the Coordination Centre but [F01] [ZGY] it on. [F01] I know. [F02] It just does my in summat chronic. [F03] Yeah. [ZGY] more [F02] Eh if I do I'll be out of my [M01] I'll get some more [F02]</pre>

http://calper.la.psu.edu/corpus_portal/tutorial_classroom.php?page=ddl



Both have pros & cons

(Boulton 2010; Frankenberg-Garcia 2012, 2016)

Hands-on





Pros ©

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- Learners can explore whatever aspect of language they want
- Serendipity: learners can learn new things they were not even looking for
- Learner autonomy

Cons 🕾

- Computers not always available in the language classroom
- Corpus software not particularly user-friendly
- Learners not always aware of what they need help with
- Learners don't always know how to build relevant/appropriate corpus queries
- Concordances can be too difficult, non-standard or inappropriate
- Learners can misinterpret the data
- Serendipity: learners can get distracted by the data

Hands-off



Pros ©

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- No computers needed in the classroom
- Learners do not have to learn to use corpus software
- Teachers aware of learner needs, can prepare tailor-made exercises
- Teachers can edit and adapt raw corpus data if too difficult, nonstandard or inappropriate

Cons 🕾

- Teachers need to learn to use corpora themselves
- Preparing own materials can be time-consuming
- Teachers need to learn how to integrate corpora to their teaching
 - Many misguided DDL materials
- What is useful to one student may not be relevant to another one
- Teachers can only cover a limited amount of language

A third approach



- Dynamic Data-Driven Learning (DDDL)
- Develop intelligent tools where
 - Learners don't have to learn to use corpora
 - But engage with data-driven learning as and when required
 - Data is curated so that learners do not get distracted or lost in the process



www.collocaid.uk

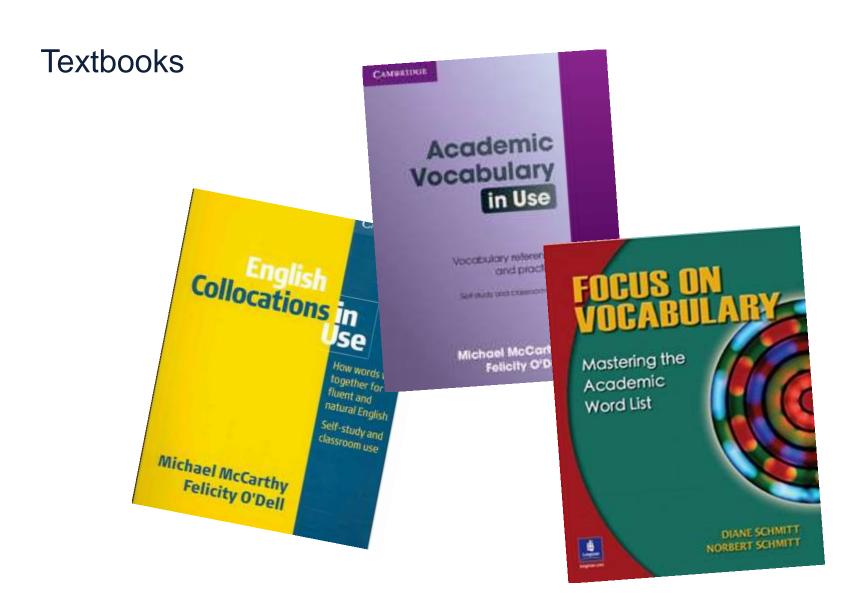
AHRC AH/P003508/1 (Surrey, Bangor and Poznan)
Ana Frankenberg-Garcia (PI), Jonathan Roberts (CI), Robert Lew (CI)
Geraint Rees (PDRA) & Nirwan Sharma (PDRA)





- A DDDL writing assistant
 - Helps writers find the words they need
- Focus on Academic English (EAP) collocations
 - The way words combine in academic English
 - carry out + research
 - statistically + significant
 - further + studies

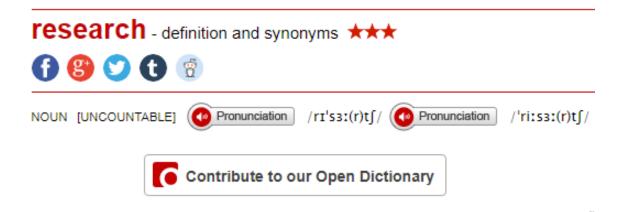






General language dictionaries

https://www.macmillandictionary.com/dictionary/british/research_1



the detailed study of something in order to discover new facts, especially in a university or scientific institution

Recent research shows that babies in the womb can be influenced by music.

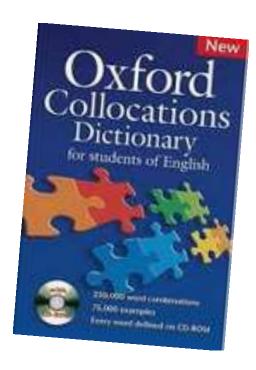
She teaches a lot of classes and doesn't have much time for her own research.

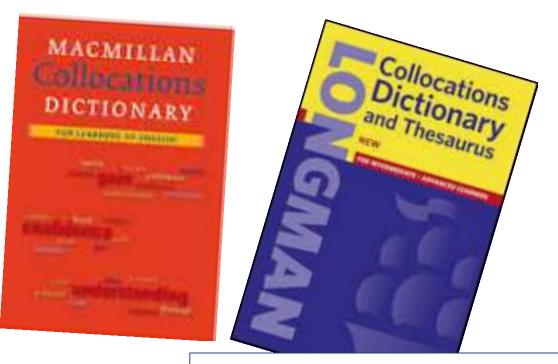
medical/historical/linguistic research

research into: research into the causes of schizophrenia research on: There is clearly a need for further research on this topic. carry out research: Government scientists have carried out extensive research into the effects of these drugs.



Collocation dictionaries

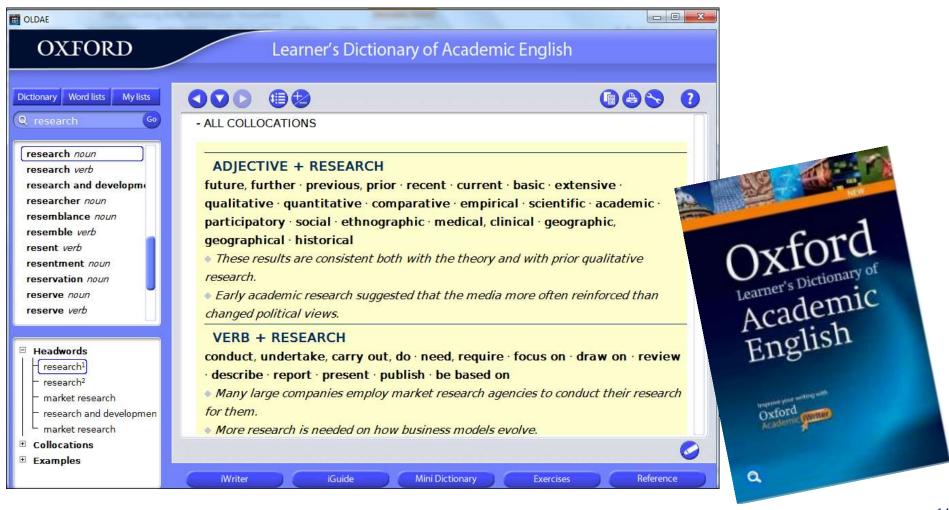




Academic collocations list (Ackermann and Chen 2013)



EAP dictionary



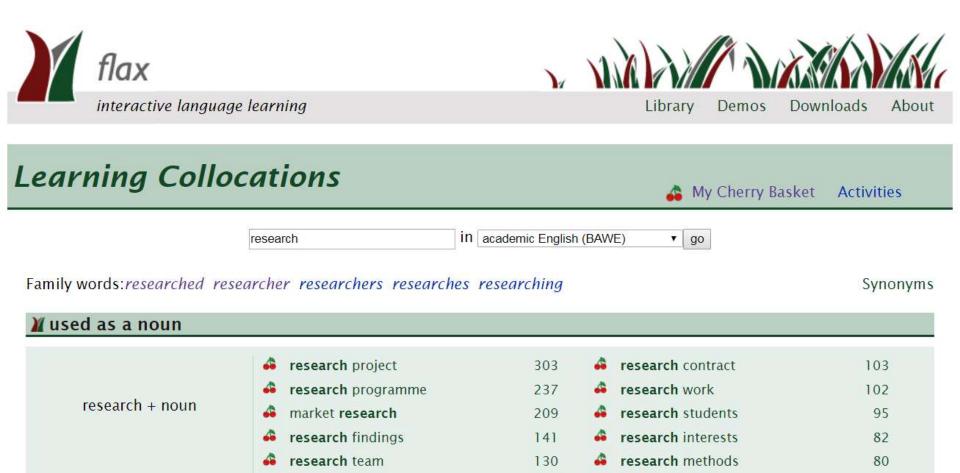


520

Corpus of Contemporary American English 👔 崖 🛂 🚱											?					
		SEARCH				FREQUE	NCY		cc	NTEXT			OVER	VIEW		
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP]													COMP	ARE		
		CONTEXT	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC Z	1990-1994 =	1995-1999 =	2000-2004 =	2005-2009	2010-2015	ALL	%	MI
1		CENTER	5789	578	78	1613	1490	2030	815	922	1181	1217	1654	153919	7.52	4.85
2		DEVELOPMENT	3871	341	47	710	672	2101	851	821	697	631	871	111382	6.95	4.74
3		INSTITUTE	3825	280	59	1274	899	1313	659	611	691	722	1142	39673	19.28	6.21
4		FUTURE	3775	28	7	101	29	3610	374	399	731	959	1312	115874	6.52	4.65
5		EDUCATION	3434	74	4	224	233	2899	375	411	465	584	1599	165391	4.15	4.00
6		UNIVERSITY	2800	171	46	772	556	1255	400	456	547	576	821	171482	3.27	3.65
7		NATIONAL	2730	171	12	630	463	1454	444	432	487	525	842	216481	2.52	3.28
8		PREVIOUS	2590	9	2	69	32	2478	265	288	495	692	850	44127	11.74	5.50
9		RESEARCH	2484	169	22	243	218	1832	347	380	413	556	788	173611	2.86	3.46
10		STUDY	2457	74	18	213	151	2001	244	270	431	526	986	177856	2.76	3.41
11		CONDUCTED	2246	49	17	291	170	1719	296	359	473	539	579	28883	15.55	5.90
									8							

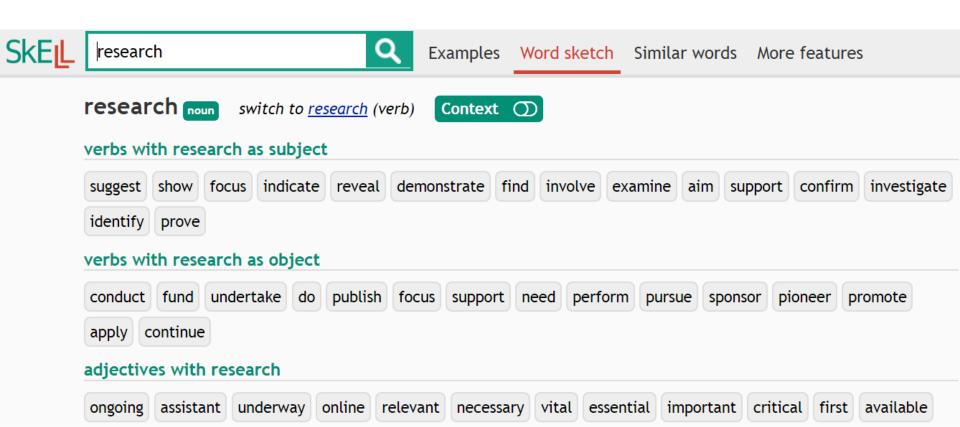
Collocates of *research* in COCA https://corpus.byu.edu/coca/ (Davies 2008)





http://flax.nzdl.org/greenstone3/flax?a=fp&sa=collAbout&c=collocations





http://skell.sketchengine.co.uk/run.cgi/skell (Baisa & Suchomel 2014)



Excellent resources available

But some limitations

- If curated (e.g. dictionaries), limited coverage
- If not curated (e.g. corpora), noise
- Learners will not look up collocations if they are not aware of their shortcomings
- Learners have to stop writing to use collocation resources (which may impair word flow)





- Will help EAP users with collocations in real time (Frankenberg-Garcia et al. 2019; Lew et al. 2018)
- Raising awareness of collocation problems

A small increase of temperature

X Do not say **increase of** something √ Say **increase in** something





- Not just correcting reactively, but helping to learn
- Raising awareness of collocations EAP users may not remember to look up in the first place

Research done as a result of...

- conduct research +
- reservab quagasta !
- emp conduct research +
- rese carry out research
 - undertake research +
- research carried out in the 1990s identified...
- before any **research** is **carried out**, it is essential that...
- this may mean that **research** can only be **carried out** in the dry season





- Customizable
- Option to switch off real-time help and check text at the end

```
A small increase of temperature...
Research done as a result of...
```

Useful for revising and proofreading





Customizable in more than one way

F	Research	one as a	re	sult of
		research + suggests + l research +		
	-	on something	+ R	estore hidden prompts research
				hypothesis
				analyse





- Metalanguage kept to a minimum
 - Users don't have to know or decypher grammatical terminology
- Data for data-driven learning will be curated so that
 - Unlike corpora, users don't get distracted from their writing with
 - irrelevant or misleading information
 - an overload of information
 - Unlike dictionaries, just the right amount
 - layered approach





Starting point are individual words, not collocations

- Learners ask "What verb can I use with research"?
- Not "Where in my text can I fit in conduct research?"

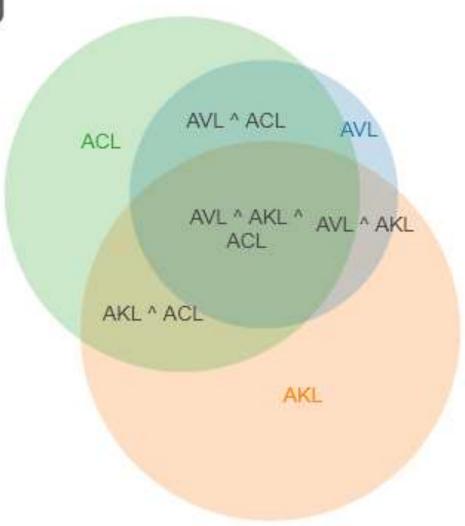
Circa 500 maximally useful interdisciplinary academic words

- 1. Academic Vocabulary List (AVL) (Gardner & Davies 2014)
 - Cross-disciplinary AVL use in BAWE (Durrant 2016)
- 2. Academic Collocations List (Ackermann & Chen 2013)
- 3. Academic Keyword List (Paquot 2010)

Methdology: 513 collocation nodes







Methodology: which collocates?





Noun bases

Adjectives & modifiers: e.g. a <u>complex</u> system, a <u>classification</u> system, an <u>existing</u> system, etc.

Verbs (object of): e.g., <u>design</u> a system, <u>develop</u> a system <u>establish</u> a system, etc.

Verbs (subject of): e.g. the system <u>consists</u> of, the system <u>functions</u> the system <u>provides</u>, etc.

Prepositions: e.g. <u>within</u> a system, a system <u>for</u> something

Verb bases

Adverbs:

e.g. <u>undoubtedly</u> contribute, contribute <u>significantly</u>, contribute <u>substantially</u>, etc.

Prepositions: e.g. contribute to

Adjective bases

Adverbs:

e.g. <u>statistically</u> significant, <u>particularly</u> significant, etc.

Methodology: which sources?





Expert academic English corpora

- Oxford Corpus of Academic English
- Pearson Corpus of Academic English
- BNC academic
- COCA academic

Methodology: which tools?





- Sketch Engine (Kilgariff et al. 2004, 2014)
- Word Sketches

X mod Nº	Y obj N°			H* subi V				
		58.20			11.14			6.40
qualitative +	1,522	10.51	conduct +	810	10.93	show +	406	9.71
future +	1.203	9.89	undertake +	192	9.37	suggest +	280	9.53
quantitative +	751	9.53	need +	278	9.05	focus +	131	9.50
further +	873	9.22	do +	357	9.01	examine	80	9.06
previous +	693	9.04	focus	84	8.33	indicate +	126	8.94
empirical +	523	8.91	carry	99	7.94	demonstrate	72	8.60
recent +	544	8.52	guide	50	7.78	explore	45	8.44
market +	472	8.38	fund	43	7.75	reveal	59	8.35
social +	845	8.24	review	61	7.75	investigate	33	8.13
marketing +	335	8.19	commission	38	7.70	address	41	8.13
survey +	258	8.03	support +	105	7.58	identify	55	8.12
scientific +	255	7.85	publish	58	7.48	find	76	7.90
action +	197	7.72	stimulate	45	7.45	involve	45	7.47
more +	354	7.65	base +	111	7.33	document	19	7.47
much +	213	7.58	summarize	37	7.23	support	39	7.44
prior +	184	7.55	warrant	28	7.17	confirm	22	7.30
current +	270	7.51	present	78	7.09	help	30	7.20
academic +	194	7.50	extend	42	7.09	highlight	18	7.14
ethnographic +	159	7.46	drive	37	6.93	entail	17	7.12
science +	162	7.45	continue	26	6.88	use	65	7.01
little +	216	7.40	report	54	6.85	need	42	7.01
basic +	185	7.18	require +	123	6.82	tend	33	6.96
health +	206	7.12	pursue	31	6.79	inform	16	6.88
extensive +	145	7.11	design	38	6.78	benefit	16	6.88
medical +	173	7.09	advance	24	6.74	concentrate	13	6.88

Methodology: which criteria?





- logDice score of ≥ 5
- co-occurrence
- ≥ 10 for lexical collocates
- ≥ 100 for prepositions
- Ignore collocates that are too general
- Ignore discipline-specific collocates

X mod N*			Y obl Nº			H* subi V				
		58.20			11.14			6.40		
qualitative +	1,522	10.51	conduct +	810	10.93	show +	406	9.71		
future +	1,203	9.89	undertake +	192	9.37	suggest +	280	9.53		
quantitative +	751	9.53	need +	278	9.05	focus +	131	9.50		
further +	873	9.22	do +	357	9.01	examine	80	9.06		
previous +	693	9.04	focus	84	8.33	indicate +	126	8.94		
empirical +	523	8.91	carry	99	7.94	demonstrate	72	8.60		
recent +	544	8.52	guide	50	7.78	explore	45	8.44		
market +	472	8.38	fund	43	7.75	reveal	59	8.35		
social +	845	8.24	review	61	7.75	investigate	33	8.13		
marketing +	335	8.19	commission	38	7.70	address	41	8.13		
survey +	258	8.03	support +	105	7.58	identify	55	8.12		
scientific +	255	7.85	publish	58	7.48	find	76	7.90		
action +	197	7.72	stimulate	45	7.45	involve	45	7,47		
more +	354	7.65	base +	111	7.33	document	19	7.47		
much +	213	7.58	summarize	37	7.23	support	39	7.44		
prior +	184	7.55	warrant	28	7.17	confirm	22	7.30		
current +	270	7.51	present	78	7.09	help	30	7.20		
academic +	194	7.50	extend	47	7.09	highlight	18	7.14		
ethnographic +	159	7.46	drive	37	6.93	entail	17	7.12		
science +	162	7.45	continue	26	6.88	use	65	7.01		
little +	216	7.40	report	54	6.85	need	42	7.01		
basic +	185	7.18	require +	123	6.82	tend	33	6.96		
health +	205	7.12	pursue	31	6.79	inform	16	6.88		
extensive +	145	7.11	design	38	6.78	benefit	16	6.88		
medical +	173	7.09	advance	24	6.74	concentrate	13	6.88		





Good corpus examples

- Not too long, not too difficult
 - Save screen space, less distracting
- Three examples of each (Frankenberg-Garcia 2014)
- Collocation highlighted in examples (Dziemianko 2014)
- Colligation cues where relevant
- research carried out in the 1990s identified...
- before any research is carried out, it is essential that...
- this may mean that research can only be carried out in the dry season

Methodology: which errors?





Data from

- Textbooks, dictionaries
- Crowdsourced feedback (www.collocaid.uk)
- Learner corpora
 - Open Cambridge Learner Corpus
 - BAWE (Nesi et al.)
 - BrAWE (Goulart da Silva 2016)
 - Overuse: create system
 - Underuse: design system





Integration with text editors

Visualisation and graphical techniques to help writers in a maximally helpful and minimally intrusive way (Roberts et al. 2017)

Hands-on





Pros ©

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- Learners can explore whatever aspect of language they want
- Serendipity: learners can learn new things they were not even looking for
- Learner autonomy

Cons 😊

- Computers not always available in the language classroom
- Corpus software not particularly user-friendly
- Learners not always aware of what they need help with
- Learners don't always know how to build relevant/appropriate corpus queries
- Concordances can be too difficult, non-standard or inappropriate
- Learners can misinterpret the data
- Serendipity: learners can get distracted by the data

Hands-off

and point and the first point on the property of the property



Pros ©

- Access to the combined intuitions of lots of speakers together
- Authenticity: how language is really used
- Answers to questions not documented in edited language resources
- No computers needed in the classroom
- Learners do not have to learn to use corpus software
- Teachers aware of learner needs, can prepare tailor-made exercises
- Teachers can edit and adapt corpus raw data if too difficult, nonstandard or inappropriate

Cons 🙁

- Teachers need to learn to use corpora themselves
- Preparing own materials can be time-consuming
- Teachers need to learn how to integrate corpora to their teaching
 - Many misguided DDL materials
- What is useful to one student may not be relevant to another one
- Teachers can only cover a limited amount of language

Demo





References



Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247.

Baisa, V., & **Suchomel, V.** (2014). SkELL – Web Interface for English Language Learning. In *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Tribun EU, 2014, pp. 63-70.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 520 million words, 1990- present. Available online at https://corpus.byu.edu/coca [06/11/2017]

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, *47*(2), 157–177.

Dziemianko, A. (2014). On the Presentation and Placement of Collocations in Monolingual English Learners' Dictionaries: Insights into Encoding and Retention. *International Journal of Lexicography*, *27*(3), 259–279.

Frankenberg-Garcia, **A.** (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146.

Frankenberg-Garcia, A. (2016) "Corpora in the Classroom". In Hall, G. (ed.) *Routledge Handbook of English Language Teaching*. London: Routledge, 383-398.

Frankenberg-Garcia, **A.** (2017). Assessing the productive collocation repertoire of writers for the development of dedicated writing assistant tools. Poster presented at <u>eLex 2017</u>, Leiden, Holland, 19-21 September 2017.

References



Frankenberg-Garcia, A. Lew, R., Roberts, J., Rees, G. and Sharma, N. (forthcoming, 2019). Developing a writing assistant to help EAP writers with collocations in real time, *ReCALL* 31(2).

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *Teaching English as a Second Language Electronic Journal* (TESL-EJ), 8(4). 1–35.

Goulart da Silva, L. (2016) *Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English.* MA Dissertation. University of Warwick, 2016.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovvář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014). The Sketch Engine: Ten Years On, *Lexicography*, 1(1). 7-36.

Kosem, I. (2010). Designing a model for a corpus-driven dictionary of Academic English. PhD thesis. Aston University, Birmingham, UK.

Lea, D. (2014). Making a Learner's Dictionary of Academic English. In A. Abel, C. Vettori, & N. Ralli (Eds.), *Proceedings of the 16th EURALEX International Congress* (pp. 181–189). Bolzano, Italy: EURAC research.

Piaget, J. (1936). La naissance de l'intelligence chez l'enfant. Paris: Delachaux & Niestle.

Roberts, J., Frankenberg-Garcia, A., Lew, R., Rees, G., & Pereda, J. (2017). Visualisation and graphical techniques to help writers write more idiomatically. Poster presented at the IEEE Conference on Visualization (VIS), October 1-6. Pheonix, Arizona.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

New directions in DDL for EAP writers: ColloCaid



internal and external audits. Thank you for your attention and cooperation with these audits. KPMG

400% More Jackoff Magic blog, and I thank you for your attention . Sylvia, I have the grave misfortune of

she is valued by her clients. $\langle p \rangle \langle p \rangle$ Thank you for your attention $| \langle p \rangle \langle p \rangle$ Gail S. Carmel, IN $\langle p \rangle \langle p \rangle$ De prepared for our global responsibility. Thank you for your attention to this urgent matter." Since op

on and medicine that works. Thank you for your attention . Cassie Travaini, RSHom, CCH is igher standard for trapping. Thank you for your attention in this matter and it really needs attentionbandon its decades old blackout rules. Thank you for your attention to this matter. Congresswoman

use would be most welcome. $\langle p \rangle \langle p \rangle$ Thank you for your attention to this vital issue, Joan, Nathalie and rcumstances. I would like to thank you for your attention to this critical issue and I welcome your

affect the length of a year? Thank you for your attention and any information will be greatly appre

ech, like his candidacy, is refreshing. Thank you for your attention to this matter. Why does lexings, for believing in the army and navy. Thank you for your attention , respectfully, Laredo Girl ... ZZZZ." The

n comes. It's that simple. Thank you for your attention to detail and great writing style. Your rts in magazines.... [groan] Thank you for your attention . [stepping off the (unscented) soap boxc Listens to the customers Innovative **Thank you for your attention** Customer's interview Customers' intervie

next year's budget process. Thank you for your attention to these issues. I look forward to your

tes. It's been a great ride, I thank you for your attention and interest. The next few years should

nce in the meantime. While thanking you for your attention , I remain. Sincerely yours,

> WASHINGTON I feel myself going. I **thank you for your attention** . You had better not take any more trou

& job seekers (of all ages). $\langle p \rangle \langle p \rangle$ Thank you for your attention and for being part of the Goliath Jobs far