

Bringing Collocation Prompts to Text Editors

Ana Frankenberg-Garcia

a.frankenberg-garcia@surrey.ac.uk



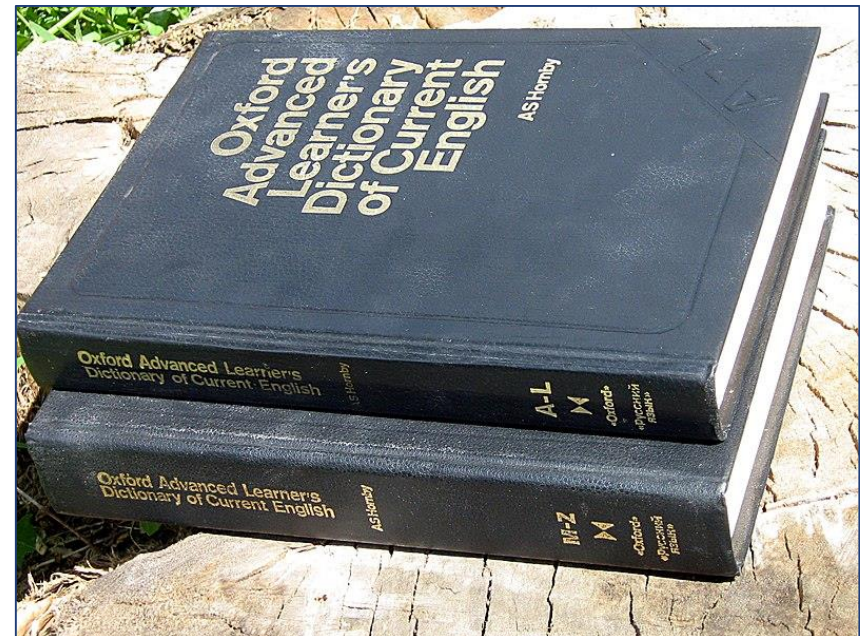
Dictionaries in the past

- Language inventory
- Word meaning
- Spelling standardization
- Etymology
- Up to users to decipher “cryptic lexicographic content”
(Lew & de Schryver 2014:341)



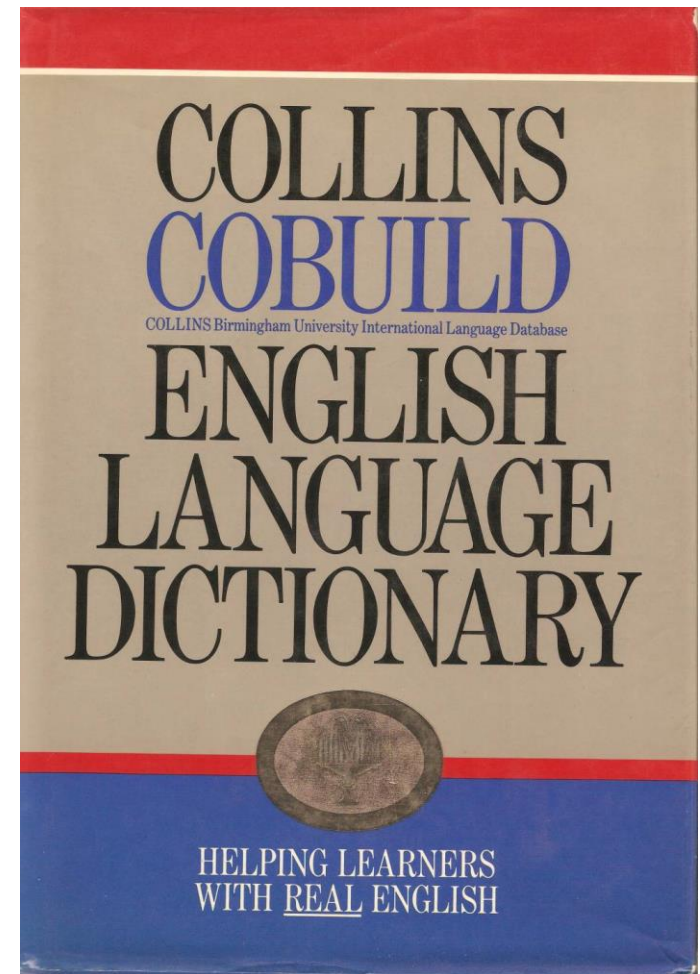
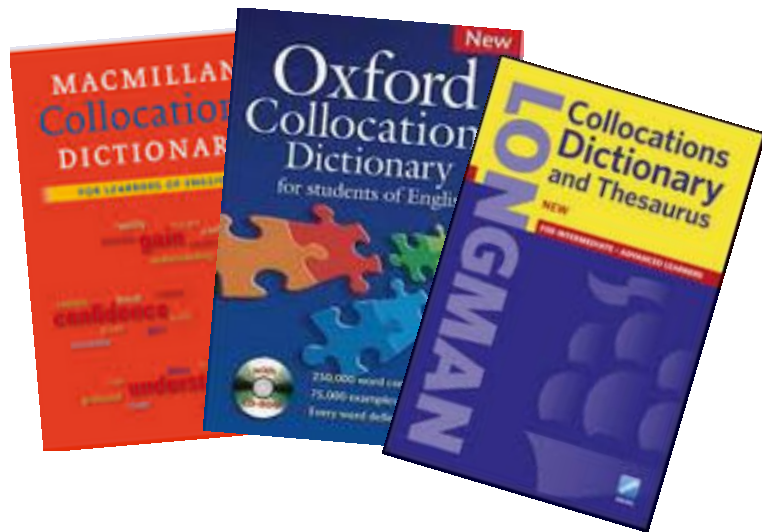
The user paradigm

- Dictionaries for learners
- Headword list selection
- Defining vocabulary
- Usage labels
- Usage patterns
- Pronunciation



The corpus paradigm

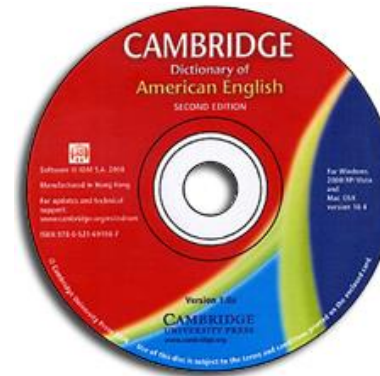
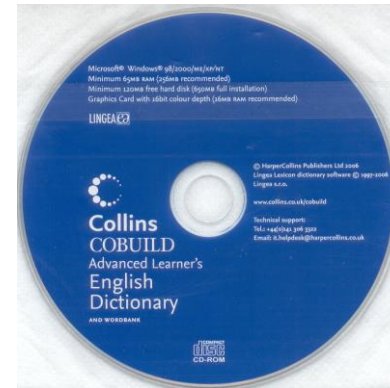
- Huge amounts of empirical data
- Word frequency
- Collocation
- Phraseology
- Authentic examples



- New types of dictionaries

The electronic medium paradigm

- Free from AZ order
- Spelling cues
- Inflected forms
- Sound files
- Enhanced visuals
- More space
 - More content
 - Content unpacked
 - More readable
- Additional resources
 - Exercises
 - Games



The portability paradigm

- Dictionaries anywhere
- Regular updates
- User contribution
- User logs



And yet...

- Spectacular developments have not had a dramatic impact on actual dictionary-user behaviour
- Paper-based and digital dictionaries remain by and large underused



- Meanings
- (Spellings)
- Translations
- Games
 - crosswords
- Authority



Language users in general are not fully aware of

- What else current dictionaries can offer them
 - e.g. most users didn't know they could use dictionaries look up collocations (Frankenberg-Garcia 2011)
- Their actual language needs
 - *e.g. the queries posed by the students suggested that they felt all they needed to become successful writers of English was a bilingual dictionary and a spelling checker* (Frankenberg-Garcia 1999:104)

- Teach more dictionary consultation skills
 - Practice with dictionary workbooks (Stark 1990)
- Raise more awareness to dictionaries
 - Integrate dictionary training in the curriculum (Carduner 2003, Frankenberg-Garcia 2011)



We use a **dictionary** to

- ❖ find the meanings of words.
- ❖ find out the correct spelling of words.
- ❖ find out how to pronounce or say words.
- ❖ find out what part of speech a word is.



<https://www.slideshare.net/acw1217/3rd-dictionary-guide-words> [15/10/2018]

<https://www.wikihow.com/Use-a-Dictionary> [15/10/2018]

1. Choose the right dictionary
2. Read the introduction
3. Learn the abbreviations
4. Learn the guide to pronunciation
5. Find the section of your dictionary with the first letter of your word
6. Read the guide words (e.g. braid-bread)
7. Scan down the page for your word
8. Read the definition
9. Alternately, you could use an online dictionary

Dictionary-like tools that

- Do not require much training or instructions
- Raise language awareness at the same time



collocaid

www.collocaid.uk



www.collocaid.uk

- UK Arts and Humanities Research Council (AH/P003508/1)
- Universities of Surrey, Bangor and Poznan
- Ana Frankenberg-Garcia (PI), Jonathan Roberts (CI), Robert Lew (CI) Geraint Rees (PDRA) & Nirwan Sharma (PDRA)





- A text editor that brings collocation help to writers
- Focus on Academic English (EAP) collocations
 - *CARRY OUT + research*
 - *STATISTICALLY + significant*
 - *FURTHER + studies*



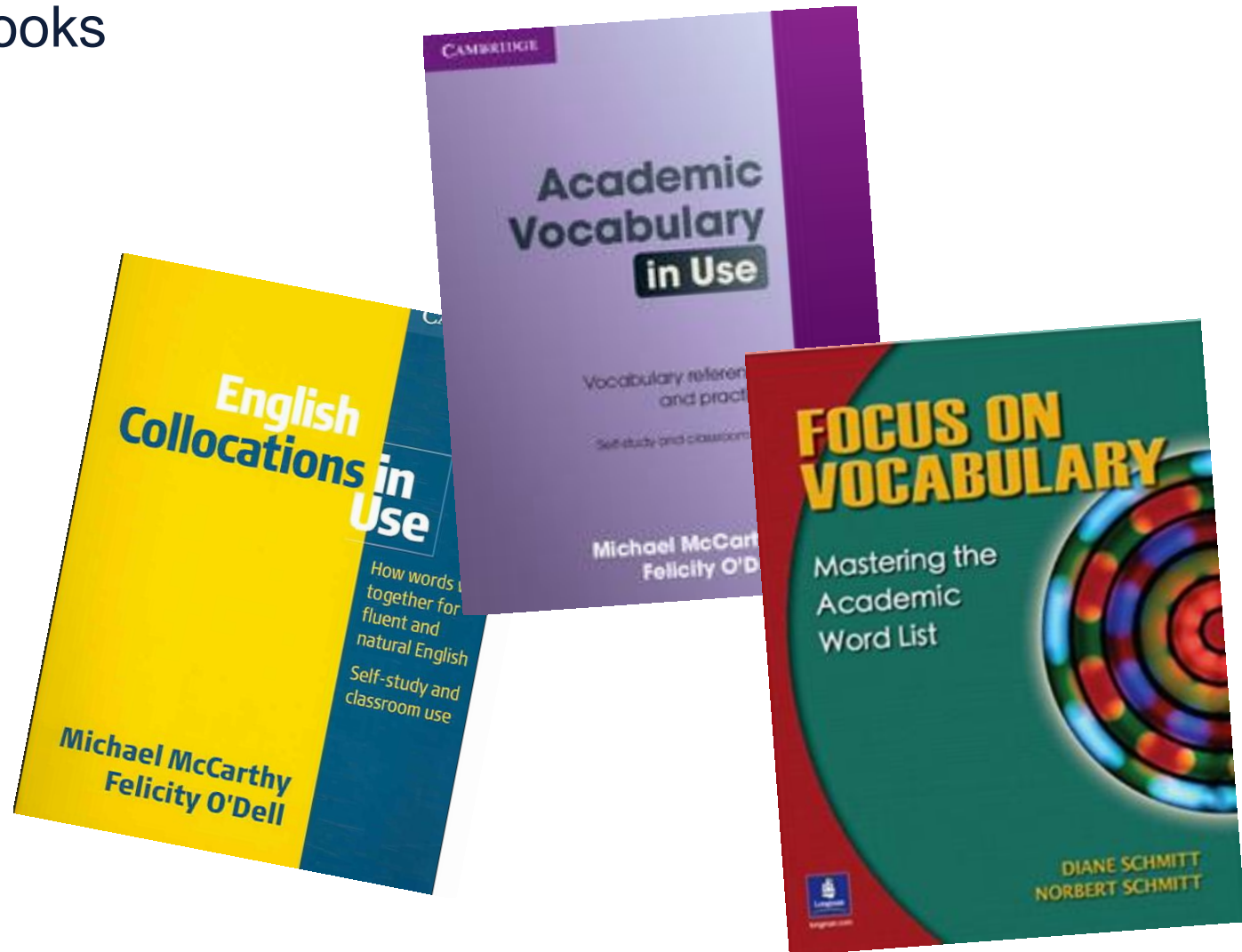
1. Collocations make texts more idiomatic, more readable (Hoey 2005, Ellis et al. 2008; Conklin and Schmitt 2012)
2. Collocations have always been an area of difficulty for language learners (Pawley and Syder 1983; Natinger and DeCarrico 1992; Howarth 1998; Wray 2002)
3. Language users often lack awareness of collocation constraints (Nesselhauf 2005; Durrant and Schmitt 2009; Laufer and Waldman 2011)



1. Not feasible to contemplate every possible collocation in a language
2. Focus on collocations for a specific group of real world users
3. Importance of English as a lingua franca for knowledge transfer (Jenkins 2014)
4. There are no native speakers of academic language (Hyland 2006, Kosem 2010, Frankenberg-Garcia 2018)

For native and non-native writers of academic English

Textbooks



General English Dictionaries for Learners

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

research - definition and synonyms ★★★



NOUN [UNCOUNTABLE]



Pronunciation

/rɪˈsɜː(r)tʃ/



Pronunciation

/'riːsɜː(r)tʃ/



Contribute to our Open Dictionary

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.

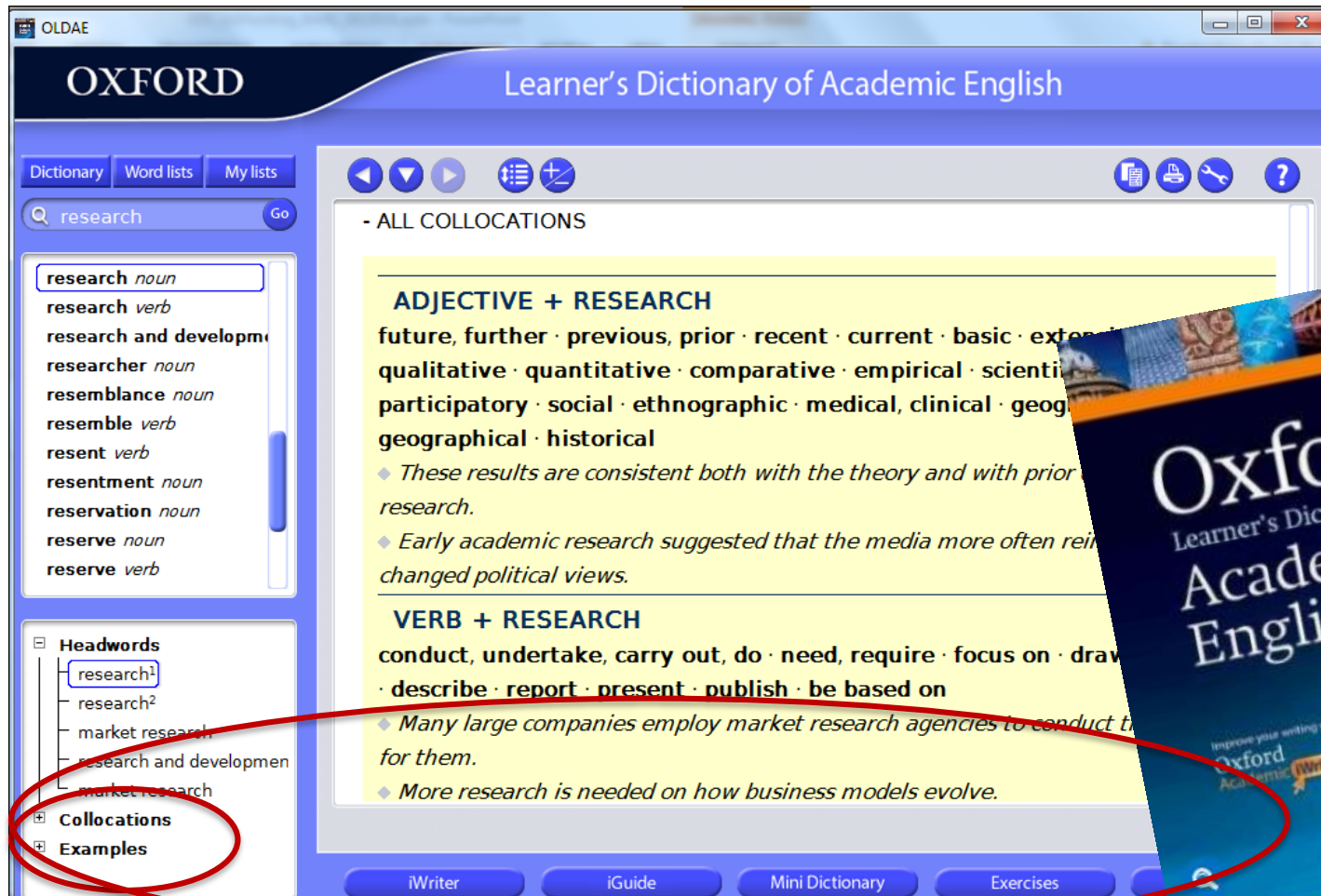
medical/historical/linguistic research

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.*

Academic English Dictionary



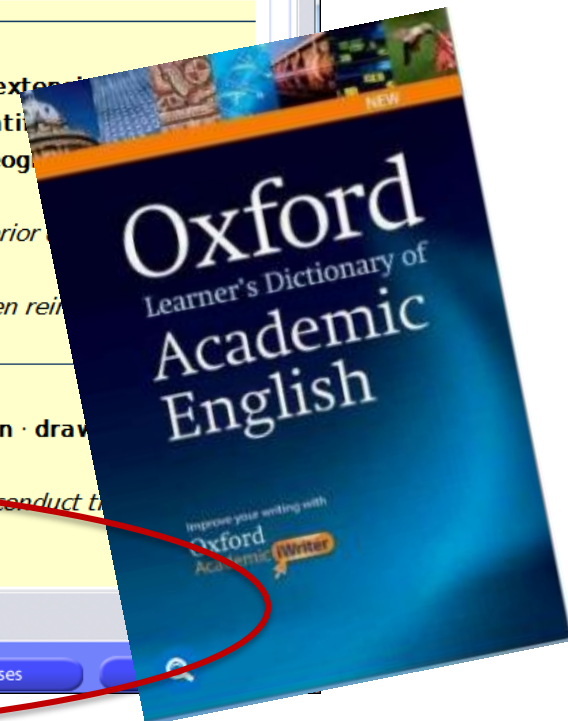
The screenshot shows the OLDAE interface with the search term 'research'. The left sidebar contains a list of words and a 'Headwords' section with a tree view. The 'Collocations' and 'Examples' sections are circled in red. The main content area displays 'ALL COLLOCATIONS' for 'research', categorized into 'ADJECTIVE + RESEARCH' and 'VERB + RESEARCH'. The 'ADJECTIVE + RESEARCH' section lists various adjectives and provides two example sentences. The 'VERB + RESEARCH' section lists various verbs and provides two example sentences.

ADJECTIVE + RESEARCH
future, further · previous, prior · recent · current · basic · extended
qualitative · quantitative · comparative · empirical · scientific
participatory · social · ethnographic · medical, clinical · geographical
geographical · historical

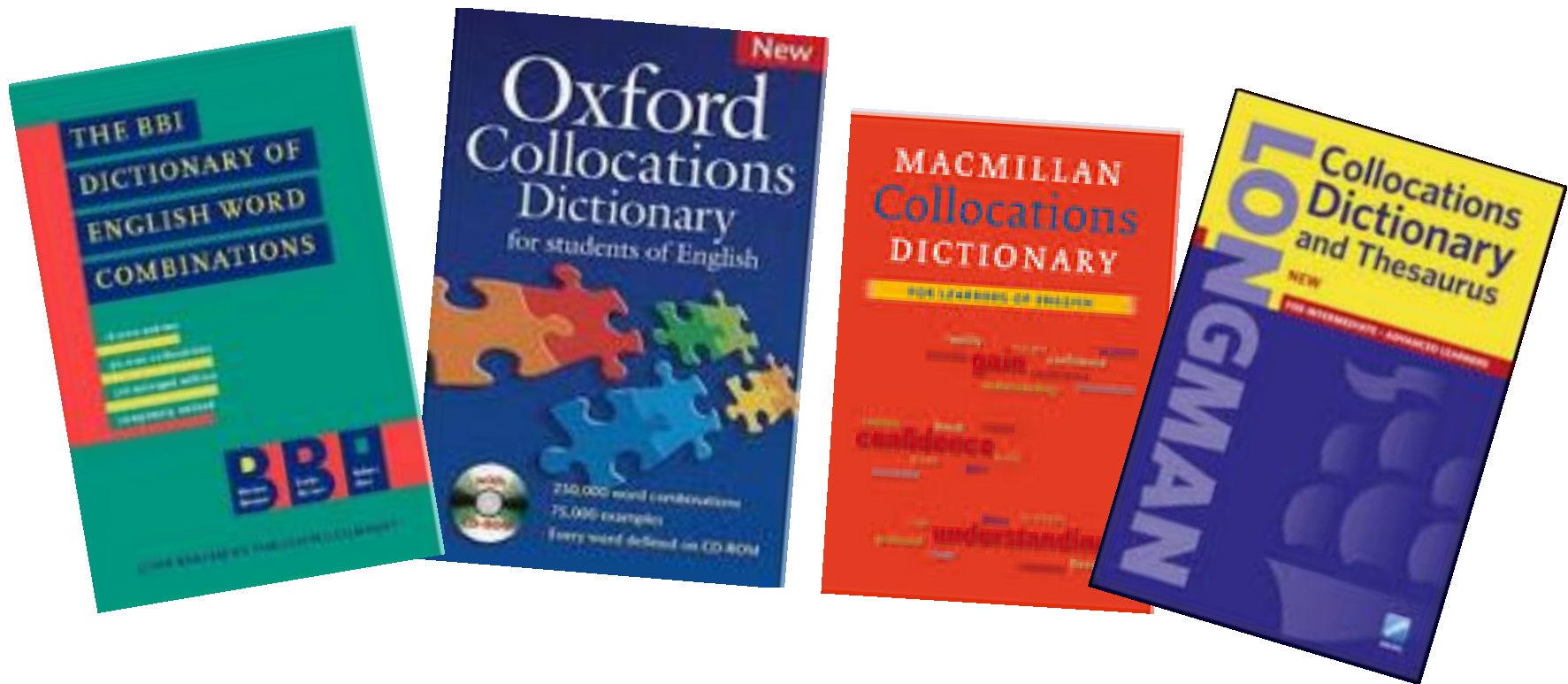
- ◆ *These results are consistent both with the theory and with prior research.*
- ◆ *Early academic research suggested that the media more often reinforced changed political views.*

VERB + RESEARCH
conduct, undertake, carry out, do · need, require · focus on · draw on
· describe · report · present · publish · be based on

- ◆ *Many large companies employ market research agencies to conduct the research for them.*
- ◆ *More research is needed on how business models evolve.*

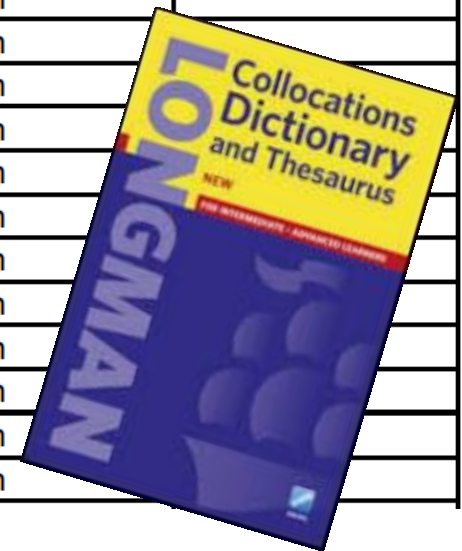


General English Collocation Dictionaries



The Academic Collocation List

#	Addition	Component I	POS I	Component II	POS II	Addition
1		abstract	adj	concept	n	
2		academic	adj	achievement	n	
3		academic	adj	career	n	
4	(in)	academic	adj	circles	n	
5		academic	adj	community	n	
6		academic	adj	debate	n	
7		academic	adj	discipline	n	
8		academic	adj	discourse	n	
9		academic	adj	institution	n	
10		academic	adj	journal	n	
11		academic	adj	life	n	
12		academic	adj	performance	n	
13		academic	adj	research	n	
14		academic	adj	skills	n	
15		academic	adj	study	n	
16		academic	adj	success	n	
17		academic	adj	work	n	



2468 “pedagogically relevant” EAP collocations (Ackermann and Chen 2013:235)

Academic Formulas List Simpson-Vlach, R., & Ellis, N. (2010)

- 1 in terms of
- 2 at the same time
- 3 from the point of view
- 4 in order to
- 5 as well as
- 6 part of the
- 7 the fact that
- 8 in other words
- 9 the point of view of
- 10 there is a
- 11 as a result of
- 12 this is a
- 13 on the basis of
- 14 a number of
- 15 there is no
- 16 point of view
- 17 the number of
- 18 the extent to which
- 19 as a result

Existing help with EAP collocations

Corpus of Contemporary American English



SEARCH

FREQUENCY

CONTEXT

OVERVIEW

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

COMPARE

	<input type="checkbox"/>	CONTEXT	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	2010-2015	ALL	%	MI
1	<input type="checkbox"/>	CENTER	5789	578	78	1613	1490	2030	815	922	1181	1217	1654	153919	7.52	4.85
2	<input type="checkbox"/>	DEVELOPMENT	3871	341	47	710	672	2101	851	821	697	631	871	111382	6.95	4.74
3	<input type="checkbox"/>	INSTITUTE	3825	280	59	1274	899	1313	659	611	691	722	1142	39673	19.28	6.21
4	<input type="checkbox"/>	FUTURE	3775	28	7	101	29	3610	374	399	731	959	1312	115874	6.52	4.65
5	<input type="checkbox"/>	EDUCATION	3434	74	4	224	233	2899	375	411	465	584	1599	165391	4.15	4.00
6	<input type="checkbox"/>	UNIVERSITY	2800	171	46	772	556	1255	400	456	547	576	821	171482	3.27	3.65
7	<input type="checkbox"/>	NATIONAL	2730	171	12	630	463	1454	444	432	487	525	842	216481	2.52	3.28
8	<input type="checkbox"/>	PREVIOUS	2590	9	2	69	32	2478	265	288	495	692	850	44127	11.74	5.50
9	<input type="checkbox"/>	RESEARCH	2484	169	22	243	218	1832	347	380	413	556	788	173611	2.86	3.46
10	<input type="checkbox"/>	STUDY	2457	74	18	213	151	2001	244	270	431	526	986	177856	2.76	3.41
11	<input type="checkbox"/>	CONDUCTED	2246	49	17	291	170	1719	296	359	473	539	579	28883	15.55	5.90
12	<input type="checkbox"/>	FURTHER	2237	36	17	173	73	1938	324	331	409	520	653	74849	5.98	4.52

Collocates of *research* in COCA

<https://corpus.byu.edu/coca/> (Davies 2008)




interactive language learning



Library Demos Downloads About

Learning Collocations











 My Cherry Basket [Activities](#)

research in academic English (BAWE)  go

Family words: *researched researcher researchers researches researching*


Synonyms

used as a noun

research + noun	 research project	303	 research contract	103
	 research programme	237	 research work	102
	 market research	209	 research students	95
	 research findings	141	 research interests	82
	 research team	130	 research methods	80

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

Learning Collocations





 My Cherry Basket [Activities](#)


research in contemporary English (Wikipedia)


Family words: *researched researcher researchers researches researching*

Synonyms

used as a noun

research + noun	 research projects 2117	 research group 1100
	 research projects 2117	 research work 1040
adjective + research	<ul style="list-style-type: none"> ACT-R, a cognitive architecture used in a large number of research projects. (A few research projects are developing CPUs based on asynchronous circuits.) One of Steinmetz's great research projects was centered with the phenomena of lightning. Complete race history of all Brabham F1 models from 1966 to 1982 and links to Brabham research projects on other models. Students write and submit SURF proposals for research projects in collaboration with professors, and about 70 percent of applicants are awarded SURFs. BMW itself was not implicated in the documentary, and the firm has 	<ul style="list-style-type: none"> ch facility 1034 ch institutions 914 ch activities 900 >>> more nic research 702 l research 694 research 615 research 550 sal research 540

SkEL | research  Examples Word sketch Similar words More features

research **noun** *switch to [research](#) (verb)* **Context** 

verbs with research as subject

suggest show focus indicate reveal demonstrate find involve examine aim support confirm investigate
identify prove


verbs with research as object

conduct fund undertake do publish focus support need perform pursue sponsor pioneer promote
apply continue

adjectives with research

ongoing assistant underway online relevant necessary vital essential important critical first available

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

SkEL  [Examples](#) [Word sketch](#) [Similar words](#) [More features](#)

research + suggest 2.09 hits per million

- 1 Some **research suggests** moods are related to overall job satisfaction.
- 2 Additional **research suggests** the social component of ritual.
- 3 Further **research suggests** cosmic dust might have skewed the results.
- 4 Other **research suggests** that later training is associated with problems.
- 5 Some **research** has **suggested** that exercise may have antidepressant effects.
- 6 This **research suggests** independence of episodic and semantic musical memory.
- 7 Some **research suggests** alternate estimates to the accepted workforce size.
- 8 Some **research suggests** an association between personality and job satisfaction.
- 9 The same **research suggests** they may have fought barefoot.
- 10 Recent **research suggests** that this is untrue.



UCL
Université
catholique
de Louvain

The Louvain EAP dictionary

Search word:

Go!

Headwords

research (n.)

research (n.)

the detailed study of a subject in order to find out new facts or information about it

Research into the impact of the media in Britain is not extensive.

Warning!

The noun *research* is rarely used in the plural.

- *Another example concerns scientific **researches**.*
- *Another example concerns scientific **research**.*

Collocations

Adj + research

recent, further, empirical, previous, future, scientific, academic, extensive, detailed, current

V + research

conduct, carry, undertake, need, base, require, publish, review, discuss, support, fund, encourage, stimulate

research + V

show, suggest, indicate, reveal, demonstrate, support

Lexical bundles

<https://leaddico.uclouvain.be/login> (Granger & Paquot 2015)

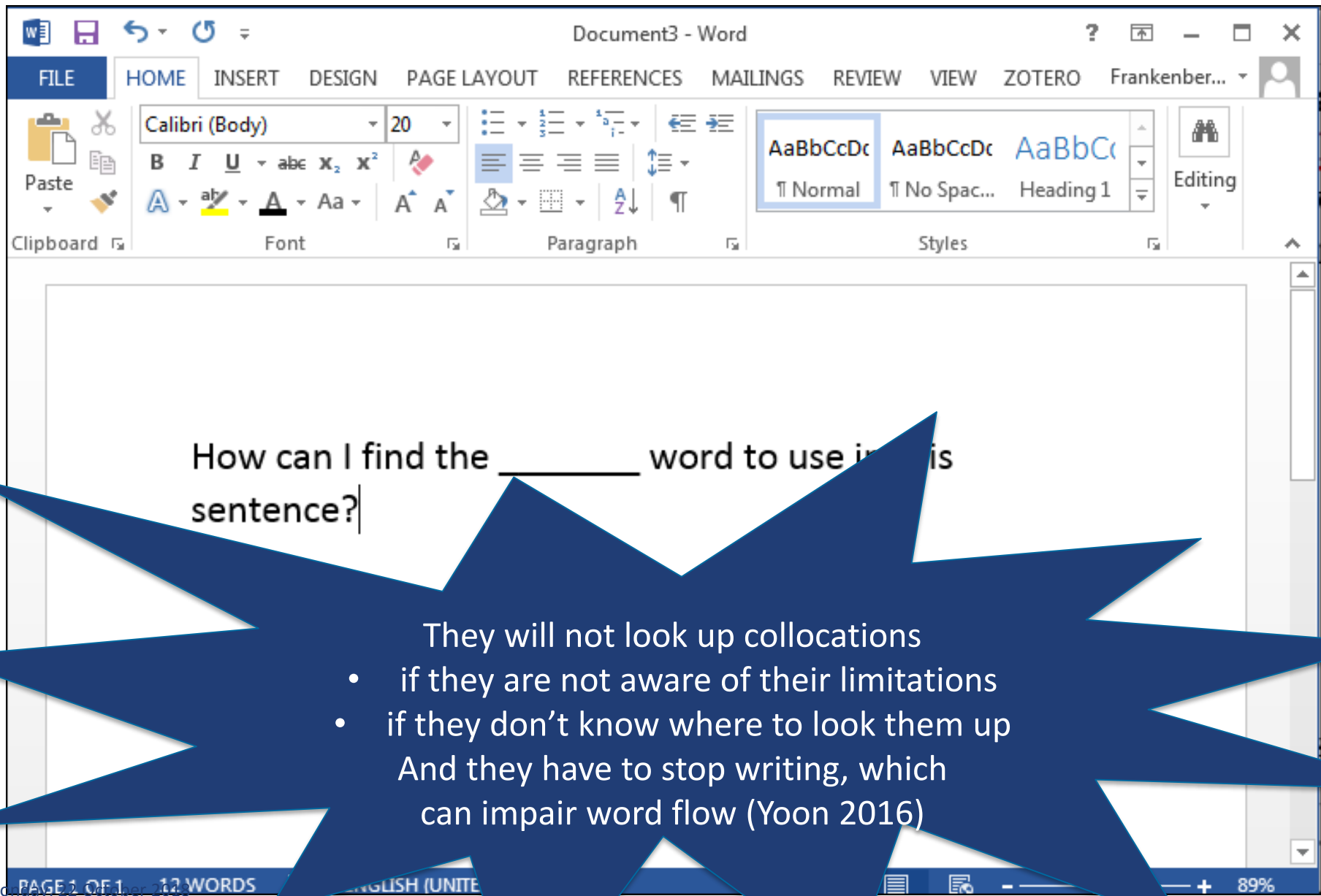
In summary

- Excellent general English resources (which include EAP)
- Excellent specific EAP resources

But some limitations

- If curated (e.g. dictionaries, textbooks), limited coverage
- If not curated (COCA, SkELL, Flax), noise
- But some combine curated and non-curated data (Louvain LEAD)

And when writers need help...



Document3 - Word

FILE HOME INSERT DESIGN PAGE LAYOUT REFERENCES MAILINGS REVIEW VIEW ZOTERO Frankenber...

Clipboard Font Paragraph Styles Editing

How can I find the _____ word to use in this sentence?

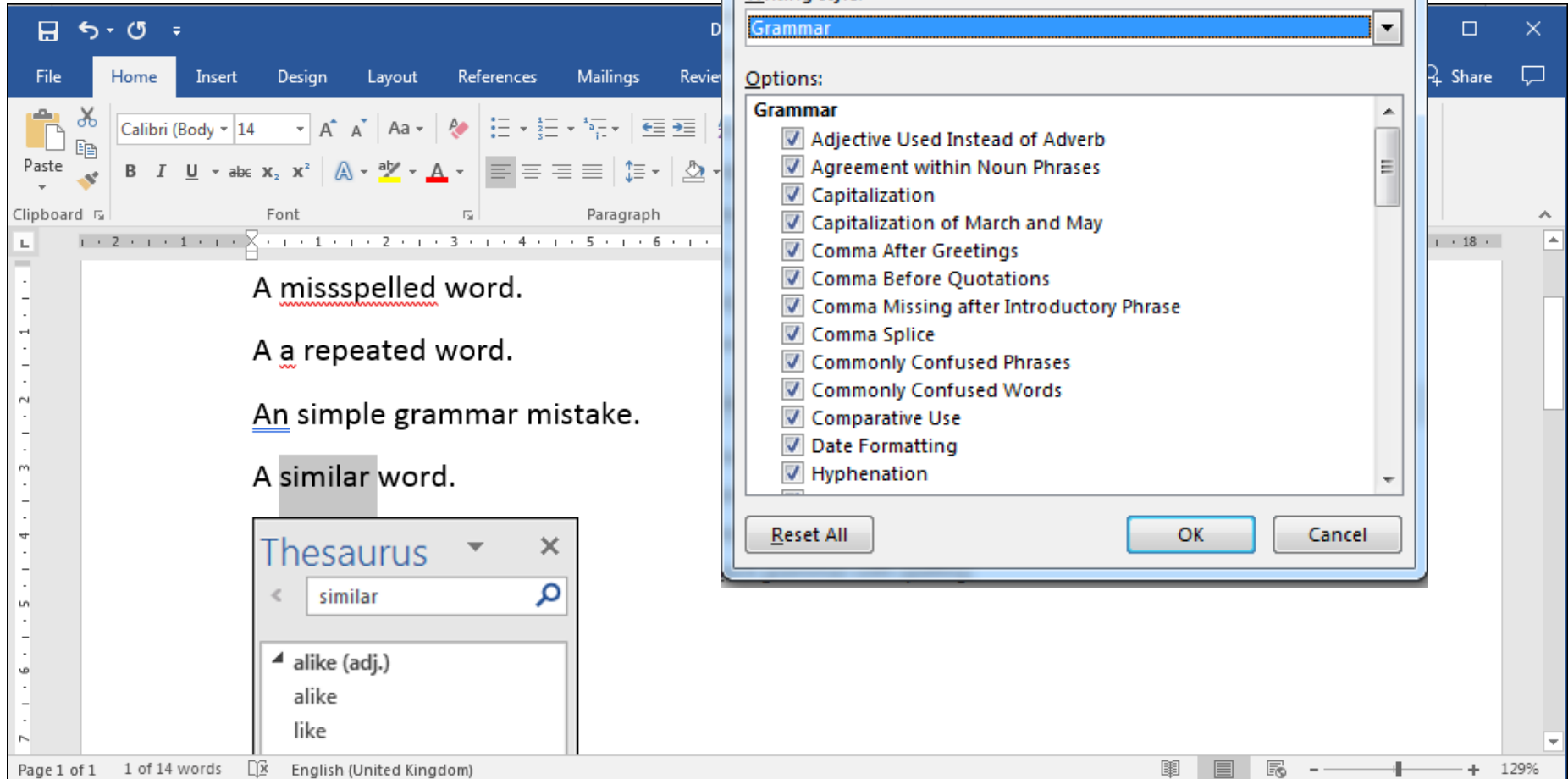
They will not look up collocations

- if they are not aware of their limitations
- if they don't know where to look them up

And they have to stop writing, which can impair word flow (Yoon 2016)

PAGE 1 OF 1 12 WORDS ENGLISH (UNITE) 89%

MS-Word



The screenshot displays the Microsoft Word interface with a document containing four lines of text, each illustrating a different type of error:

- A misspelled word. (The word "misspelled" is underlined with a red squiggly line.)
- A repeated word. (The word "a" is underlined with a red squiggly line.)
- An simple grammar mistake. (The word "An" is underlined with a red squiggly line.)
- A similar word. (The word "similar" is highlighted in grey.)

The Grammar Settings dialog box is open, showing the following options:

Writing style: Grammar

Options:

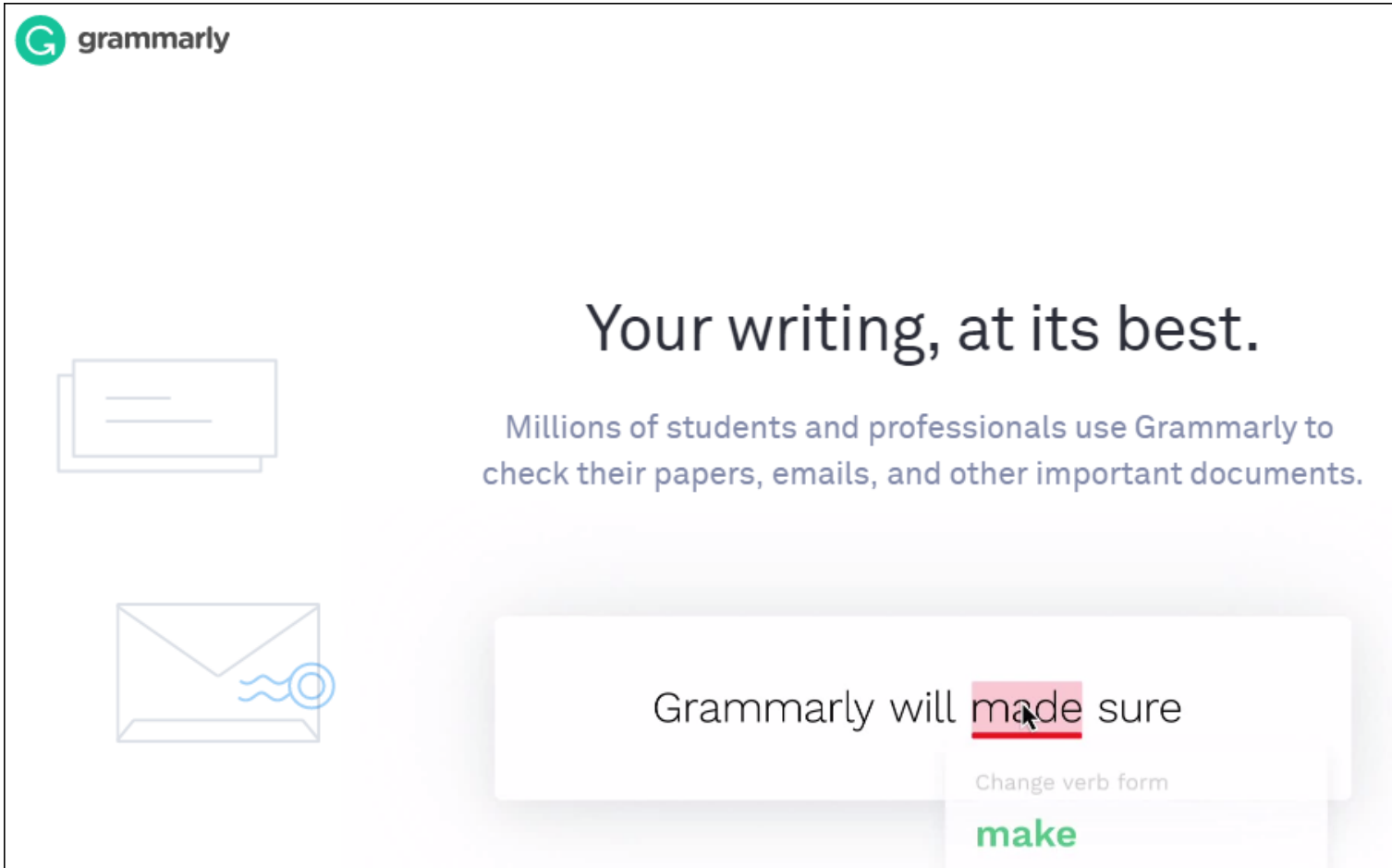
- Grammar
- Adjective Used Instead of Adverb
- Agreement within Noun Phrases
- Capitalization
- Capitalization of March and May
- Comma After Greetings
- Comma Before Quotations
- Comma Missing after Introductory Phrase
- Comma Splice
- Commonly Confused Phrases
- Commonly Confused Words
- Comparative Use
- Date Formatting
- Hyphenation

Buttons: Reset All, OK, Cancel

The Thesaurus window is also open, showing the word "similar" and its suggestions: alike (adj.), alike, and like.

Page 1 of 1 | 1 of 14 words | English (United Kingdom) | 129%

Grammarly <https://www.grammarly.com/>



The screenshot shows the Grammarly website homepage. In the top left corner is the Grammarly logo, which consists of a green circle with a white 'G' and the word 'grammarly' in lowercase. Below the logo are two icons: one of a document with lines representing text, and another of an envelope with a blue circular icon containing a checkmark. The main heading is 'Your writing, at its best.' followed by the text 'Millions of students and professionals use Grammarly to check their papers, emails, and other important documents.' At the bottom, there is a white box containing the text 'Grammarly will **made** sure'. The word 'made' is highlighted with a red background. Below this text is a dropdown menu with the option 'Change verb form' and the word 'make' in green.

grammarly

Your writing, at its best.

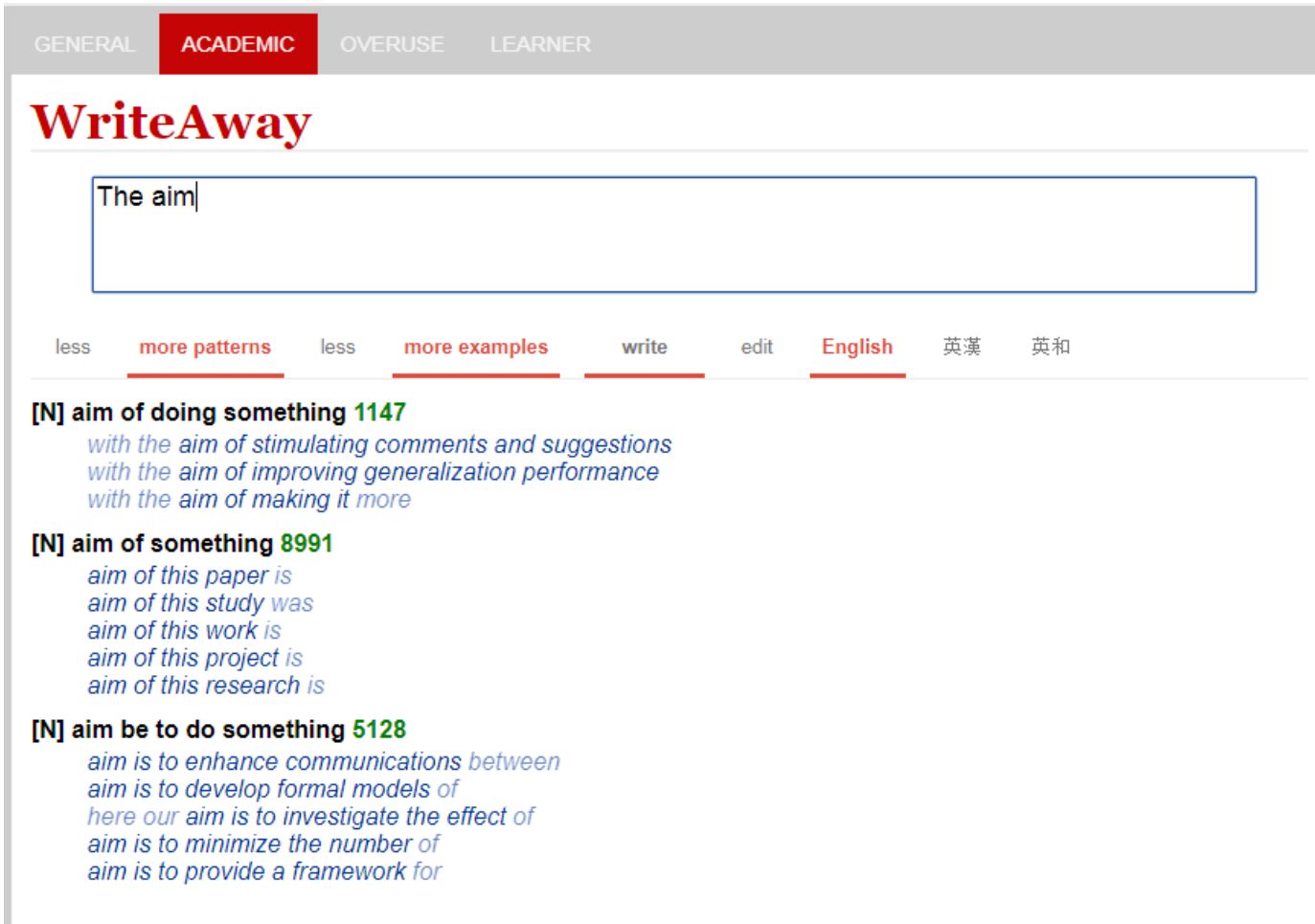
Millions of students and professionals use Grammarly to check their papers, emails, and other important documents.

Grammarly will **made** sure

Change verb form

make

WriteAway (<http://writeaway.nlpweb.org/>)



The screenshot shows the WriteAway website interface. At the top, there are navigation tabs: GENERAL, ACADEMIC (highlighted in red), OVERUSE, and LEARNER. Below the tabs is the title "WriteAway" in a large, bold, red font. A text input field contains the text "The aim". Below the input field, there are several navigation links: "less", "more patterns" (highlighted in red), "less", "more examples" (highlighted in red), "write", "edit", "English" (highlighted in red), "英漢", and "英和". The main content area displays three categories of examples for the word "aim":

- [N] aim of doing something 1147**
 - with the aim of stimulating comments and suggestions*
 - with the aim of improving generalization performance*
 - with the aim of making it more*
- [N] aim of something 8991**
 - aim of this paper is*
 - aim of this study was*
 - aim of this work is*
 - aim of this project is*
 - aim of this research is*
- [N] aim be to do something 5128**
 - aim is to enhance communications between*
 - aim is to develop formal models of*
 - here our aim is to investigate the effect of*
 - aim is to minimize the number of*
 - aim is to provide a framework for*

In summary

- Some truly innovative tools

But some limitations

- Mostly reactive feedback (except predictive text)
- Feedback can be
 - prescriptive
 - distracting
 - intrusive
 - annoying



- Our vision is to offer collocation help in real time
- Will raise awareness of collocation problems

A small increase of temperature

✗ Do not say **increase of temperature**
✓ Say **increase in temperature** (+)

- the reasons for this **increase in temperature** are not fully understood
- a temporary **increase in temperature** that can result in loss of...
- symptoms can deteriorate with an **increase in temperature**

- No auto-correction, helping to learn



- But not just helping reactively
- Raising awareness of collocations EAP users may not remember to look up in the first place
- Not just feedback, but also feed forward

Research _____ in the UK...

• conduct research (+)

• research suggests +

• empirical research +

• research on (the behaviour)

• conduct research +

• 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 in the UK..

- Useful for revising and proofreading

What's different about ColloCaid?



- Customizable in more than one way

Research carried out in the UK...

- conduct research +
- research suggests +
- empirical research +
- research on (the behaviour/act
- research in (the field/the conte

Hide

Restore hidden prompts

- research
- hypothesis
- analyse

What's different about ColloCaid?



- Works in real time, writers do not have to stop writing to consult external resources
- Is not just about feedback on miscollocations...
- Helps writers to expand their collocation repertoire autonomously
- Constructivist approach (Piaget 1936, Vygotsky 1978)
- Fosters discovering and noticing instead of rote learning
 - Data-driven learning (Johns 1991)

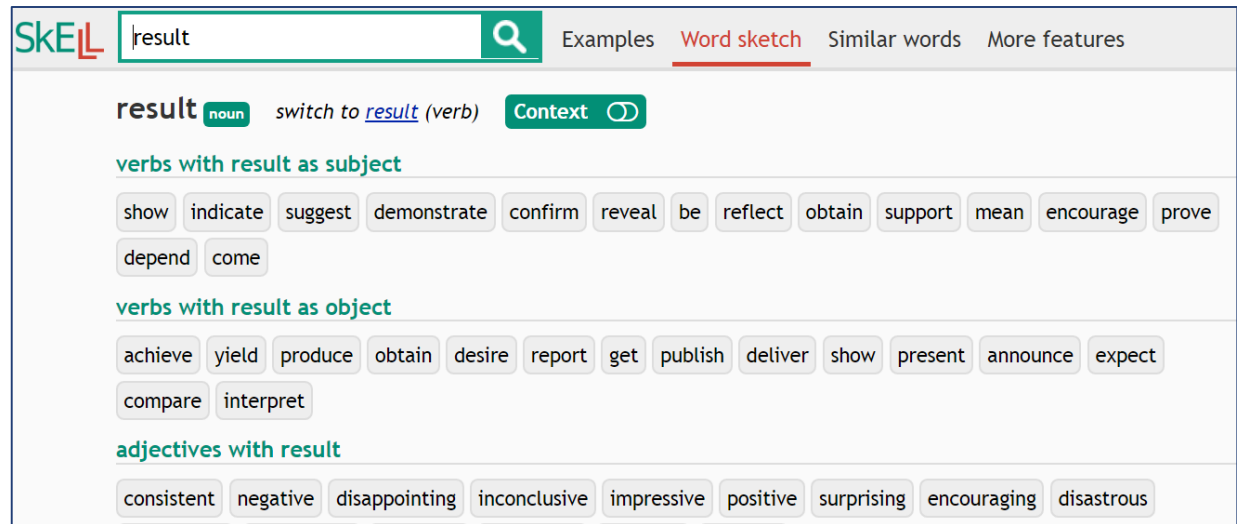


- Metalanguage is kept to a minimum
 - Users don't have to know or decypher grammatical terminology
- Lexicographic data is curated (scaffolding) so that users don't get distracted from their writing with
 - irrelevant or misleading information
 - information that is difficult to find
 - an overload of information
 - a flexible, layered approach, with as much (or as little) information as needed



- But will provide links to non-curated collocation data not covered in Collocaid

The results of the...



The screenshot shows the SKELL website interface. At the top, there is a search bar containing the word 'result' and a magnifying glass icon. To the right of the search bar are navigation links: 'Examples', 'Word sketch' (highlighted in red), 'Similar words', and 'More features'. Below the search bar, the word 'result' is displayed with a 'noun' tag and a link to 'switch to result (verb)'. There is also a 'Context' button. The main content area is divided into four sections, each with a title and a list of related words in rounded rectangular buttons:

- verbs with result as subject**: show, indicate, suggest, demonstrate, confirm, reveal, be, reflect, obtain, support, mean, encourage, prove, depend, come
- verbs with result as object**: achieve, yield, produce, obtain, desire, report, get, publish, deliver, show, present, announce, expect, compare, interpret
- adjectives with result**: consistent, negative, disappointing, inconclusive, impressive, positive, surprising, encouraging, disastrous

- User logs feed back to Collocaid database

What's different about Collocaid?

collocaid

- Usability testing will be a priority





Starting point are individual words, not collocations

- Writers ask “*What verb can I use with **research**?*”
- Not “*Where in my text can I fit in **conduct research**?*”
- Collocation nodes

Which ones?

- 500 maximally useful

Methodology: which collocation nodes?



What are the 500 maximally useful collocation nodes?

- The most frequent academic words used across disciplines
- Why most frequent?
 - Top-ranking words cover most of the language (Nation 2001)
- Why interdisciplinary?
 - People get targeted and concentrated exposure to the vocabulary used in their disciplines
 - End up learning them incidentally, without too much effort
 - Interdisciplinary academic words are less salient, so probably harder to learn



How to determine the 500 most frequent academic words used across disciplines?

- Use existing research rather than reinvent the wheel...
- Lists of most frequent words from corpora of millions of words of academic English
 - Academic Word List (Coxhead 2000)
 - Academic Keyword List (Paquot 2010)
 - Academic Collocation List (Ackermann and Chen 2013)
 - Academic Vocabulary List (Gardner & Davies 2014)
- But different corpora, different extraction methods,
 - Different lists...



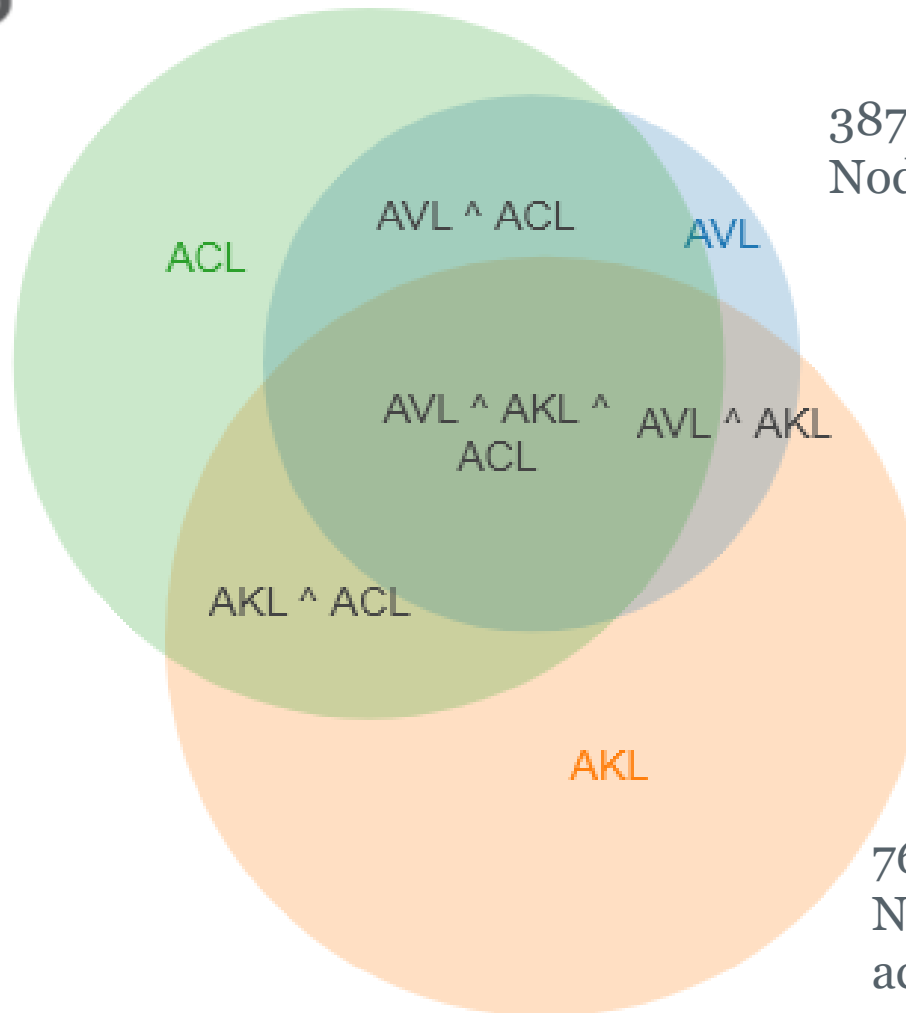
Our solution: prioritize what known lists have in common

Build on strengths of each list

- ~~Academic Word List (Coxhead 2000)~~
- Academic Keyword List (Paquot 2010)
- Academic Collocation List (Ackermann and Chen 2013)
- Academic Vocabulary List (Gardner & Davies 2014)
 - 3000 lemmas: no time & no point covering lemmas not used
 - Subset in student writing from BAWE corpus (Durrant 2016)
- Focus on nouns, verbs and adjectives
 - Nouns (e.g. activity)
 - Verbs (e.g. demonstrate)
 - Adjectives (e.g. appropriate)
- Exclude
 - Adverbs (e.g. primarily)



703 ACL lemmas:
Nodes that evoke
strong collocations



387 AVL-BAWE lemmas:
Nodes used by students

766 AKL lemmas:
Nodes that stand out in
academic texts



- Vocabulary lists tend not to cover compounds
 - e.g. *case study, research question, experimental design*
 - Covered under main word
 - *study, question, design*
- Some initial nodes not collocationally productive
 - Mostly adjectives (e.g. *alternative*)
 - Not covered in ColloCaid



- Some nodes with more than one part of speech
 - e.g. *research* > noun & verb
 - introduce disambiguation step
- Some nodes with more than one part of speech but only one is covered in selected list
 - e.g. *aid; base; construct* > verbal only
 - complement coverage with nominal uses

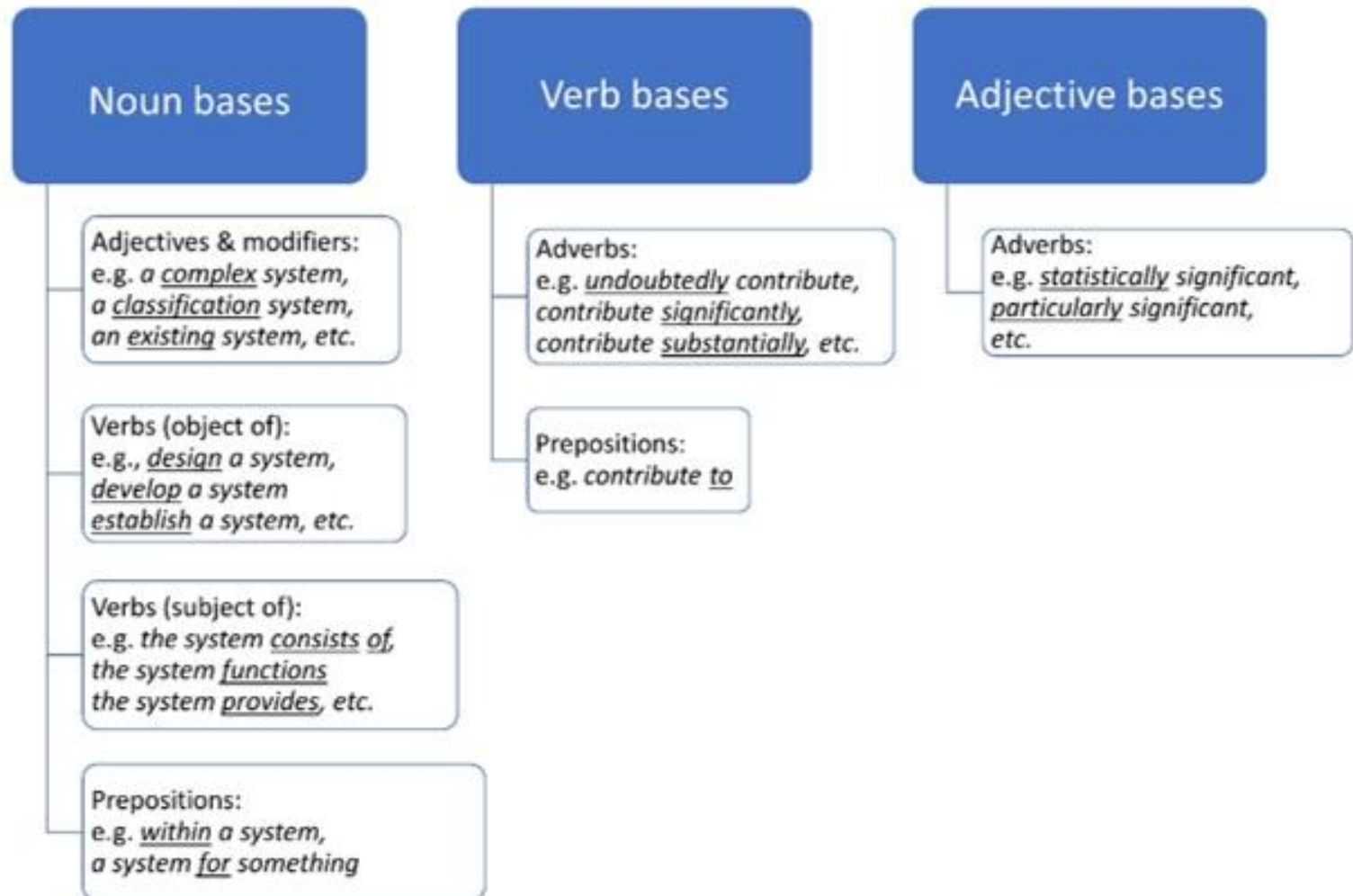


- Some nodes with more than one sense
 - cover all academic senses
 - *figure* (= as number; as image; as person)
 - *model* (=as example; ~~in fashion; in art~~)
- Some nodes not interdisciplinary, but multidisciplinary
 - e.g. *code* (= in computing; in biology; as symbol)
 - cover discipline-specific nodes used in more than one discipline
 - will pave the way for future discipline-specific customization



- Flagging up node lemmas alone not enough
- Integration with text editors means inflected forms need to be recognized
 - apply/applies/applied/applying
- And spelling variants too
 - analyse/analyze
- Tool won't recognize misspellings
 - We need to include a spell checker

Methodology: which collocates?





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

Word Sketch for *research*
in OCAE

<u>X_mod N*</u>			<u>Y_obj N*</u>			<u>N* subj Y</u>		
		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 +	306	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 collocates?



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

<u>X_mod N*</u>			<u>Y_obj N*</u>			<u>N* subj Y</u>		
		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 +	306	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



Node-collocate order

- noun + VERB: *approach allows*
- VERB + noun: *adopt approach*
- ADJECTIVE + noun: *alternative approach*
- ADVERB + verb: *successfully apply*
- verb + ADVERB : *apply equally*
- verb + PREPOSITION:
 - *increase in (speed/rate/level of)*
 - *increase of (10 per cent)*
 - *increase over (time/the period)*



Good corpus examples (Atkins & Rundell 2008)

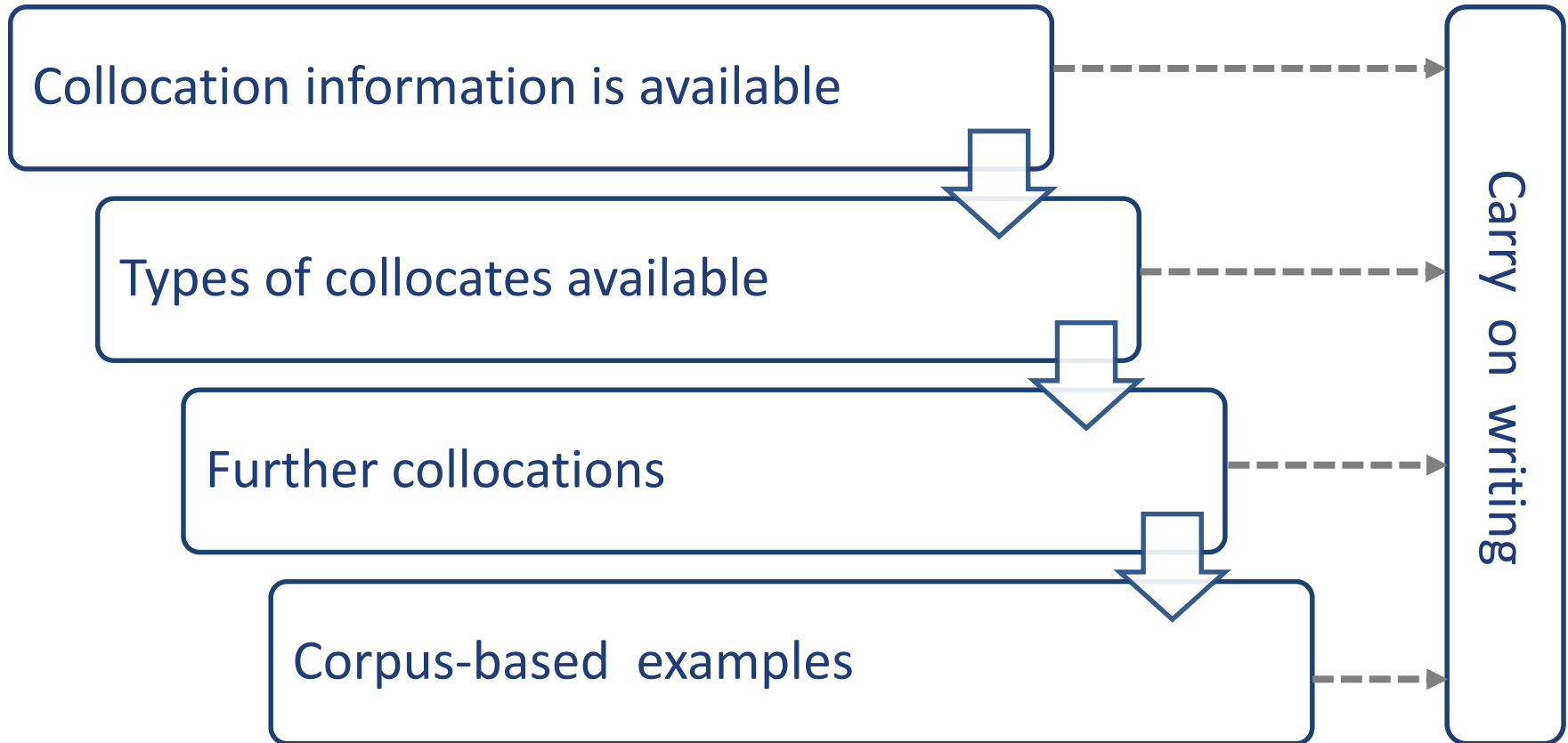
- Not too long,
 - Save screen space, less distracting
 - Fair use (no plagiarism)
- Not too difficult
 - GDEX (Kilgarriff et al. 2008)
- Anonymized
 - References in number format
- Three examples of each (Frankenberg-Garcia 2014)
 - Collocation highlighted in examples (Dziemianko 2014)
- Colligation 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



Data from

- Textbooks, dictionaries
- Crowdsourced feedback
- Learner corpora
 - BAWE (Nesi et al.)
 - Error: *based in (something)*
 - Overuse: *create system*
 - Underuse: *design system*
 - Cambridge Learner Corpus





- Link lexicographic database with text editor
- Initial prototype developed in open source online editor
 - TinyMCE (www.tiny.cloud)
- Can update versions
- Integrate feedback
 - for missing words
 - other usability issues
- User experience is key!



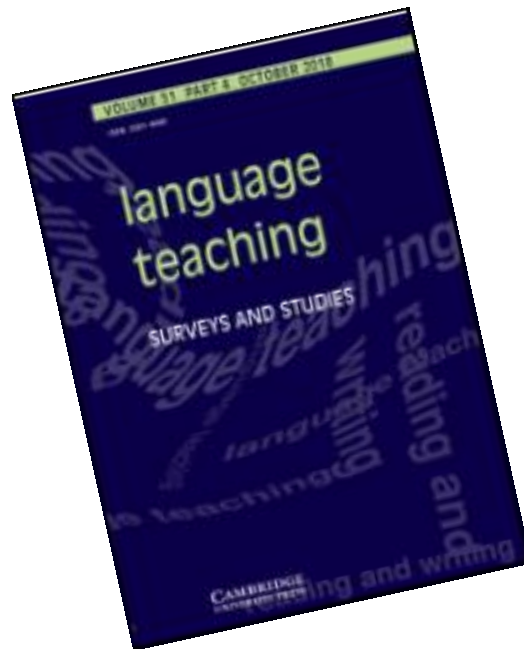
Seek expert feedback from colleagues

Test prototype with EAP users

- build a prototype +
- a prototype featuring +
- an experimental prototype +
- a prototype of (the tool) +

Frankenberg-Garcia, A. (2018) [Combining user needs, lexicographic data and digital writing environments](#). *Language Teaching*. First view online, 1-15.

Frankenberg-Garcia, A., Lew, R., Roberts, J., Rees, G. and Sharma, N. (2018) [Developing a writing assistant to help EAP writers with collocations in real time](#). *ReCALL*. First view online, 1-17.



File ▾ Edit ▾ View ▾ Insert ▾ Format ▾ Tools ▾ Table ▾ Help ▾

Paragraph ▾ **B** *I* ~~S~~ A ▾ **A** ▾      ▾  ▾  ▾   

I