

A thick black L-shaped frame surrounds the text. The top-left corner is a horizontal bar extending to the right, then a vertical bar extending downwards. The bottom-right corner is a horizontal bar extending to the right, then a vertical bar extending upwards.

ColloCaid: A real-time tool to help academic writers with English collocations

*Robert Lew, Ana Frankenberg-Garcia, Geraint Rees,
Jonathan Roberts, and Nirwan Sharma*

EURALEX 2018

Away from static dictionaries

- the *digital revolution*
- traditional stand-alone dictionaries are on their way out
- dictionaries as we know it giving way to digital tools
- geared towards specific tasks

Collocation (for Academic English)

- Appropriate collocations make texts more idiomatic, more readable (Hoey 2005, Ellis et al. 2008; Conklin and Schmitt 2012)
- Collocation has always been a particular area of difficulty for language learners (Pawley and Syder 1983; Natinger and DeCarrico 1992; Howarth 1998; Wray 2002)
- Language users often lack awareness of collocational constraints (Nesselhauf 2005; Durrant and Schmitt 2009; Laufer and Waldman 2011)
- Importance of English as a lingua franca for the dissemination of science (Jenkins 2014)
- There are no native speakers of academic language (Hyland 2006, Kosem 2010, Frankenberg-Garcia 2017)

EXISTING RESOURCES
FOR COLLOCATION:
DICTIONARIES



LDOCE Collocation box



COLLOCATIONS

VERBS

do/carry out research (also **conduct research** *formal*)

The research was carried out by a team of scientists at Edinburgh University.
Little research has been conducted into the subject.

undertake research *formal* (=start or do research)

They are planning to undertake research into the genetic causes of the disease.

ADJECTIVES

scientific research

Our conclusions are based on scientific research.

medical research

The charity raises money for medical research.

historical research

This is a fascinating piece of historical research.

basic research (=the most important or most necessary area of research)

He wants to conduct basic research into the nature of human cells.

extensive research (=research that examines a lot of information and details)

Extensive research has proved the theory wrong.

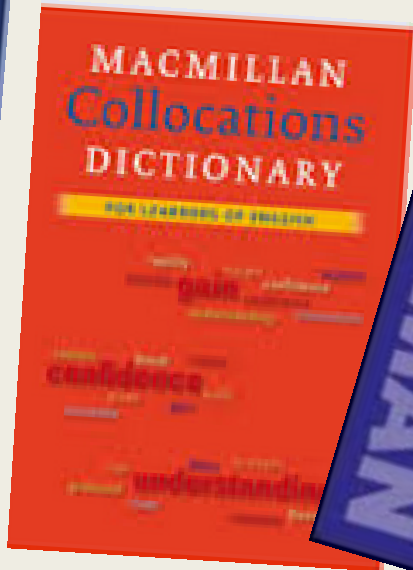
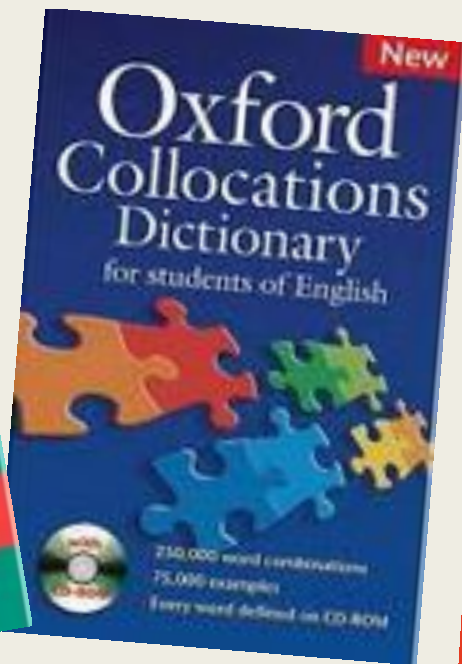
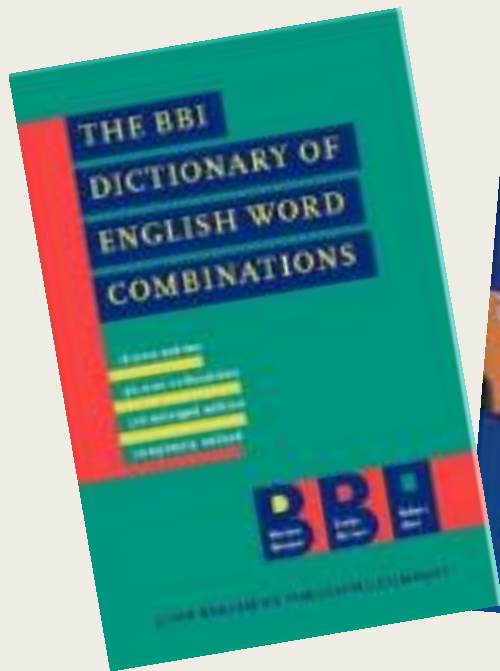
painstaking research (=very careful and thorough research)

She spent years carrying out painstaking historical research.

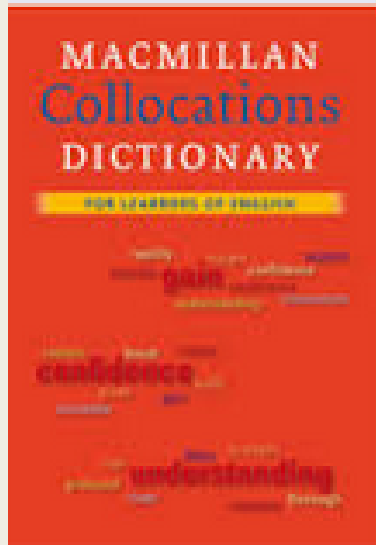
pioneering research (=research that produces completely new information)

Watson did pioneering research on the long-term effects of dieting.

Collocation Dictionaries



Macmillan Collocations Dictionary



- ▶ using particular methods **academic, applied, comparative, empirical, qualitative, quantitative** *Much of the research is qualitative and based on both male and female samples.*
- v+N do research **carry out, conduct, do, perform, pursue, undertake** *She is continuing to do research into medieval European dress.*
- ▶ provide money for research **finance, fund, sponsor, support** *Every penny received goes directly to buy equipment or to fund research.*
- ▶ officially ask someone to do research **commission** *The government has commissioned further research to be carried out by The National Parenting and Families Institute.*
- ▶ tell people the results of research **deliver, present, publish, report** *Professor David Mellor is to present his research at a conference organised by Compassion in World Farming next month.*
- ▶ be in charge of a programme of research **co-ordinate, direct, lead, manage, supervise** *The research was led by Professor Dennis Parker.*
- ▶ look at other people's research **assess, evaluate, review** *This paper will review the research of the past decade and consider some of the gaps in our knowledge.*
- ▶ report or publish your research **disseminate, present, publicize, publish, report** *The monthly meeting provides a forum for team members to present their research.*

EXISTING
RESOURCES:
CORPORA



SkELL (Baisa & Suchomel 2014)

SkELL

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

SkELL (Baisa & Suchomel 2014)

SkELL



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.



flax

interactive language learning



[Library](#) [Demos](#) [Downloads](#) [About](#)

Learning Collocations











 [My Cherry Basket](#) [Activities](#)

in


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

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"> research facility 1034 research institutions 914 research activities 900 >>> more academic research 702 biological research 694 environmental research 615 medical research 550 social research 540

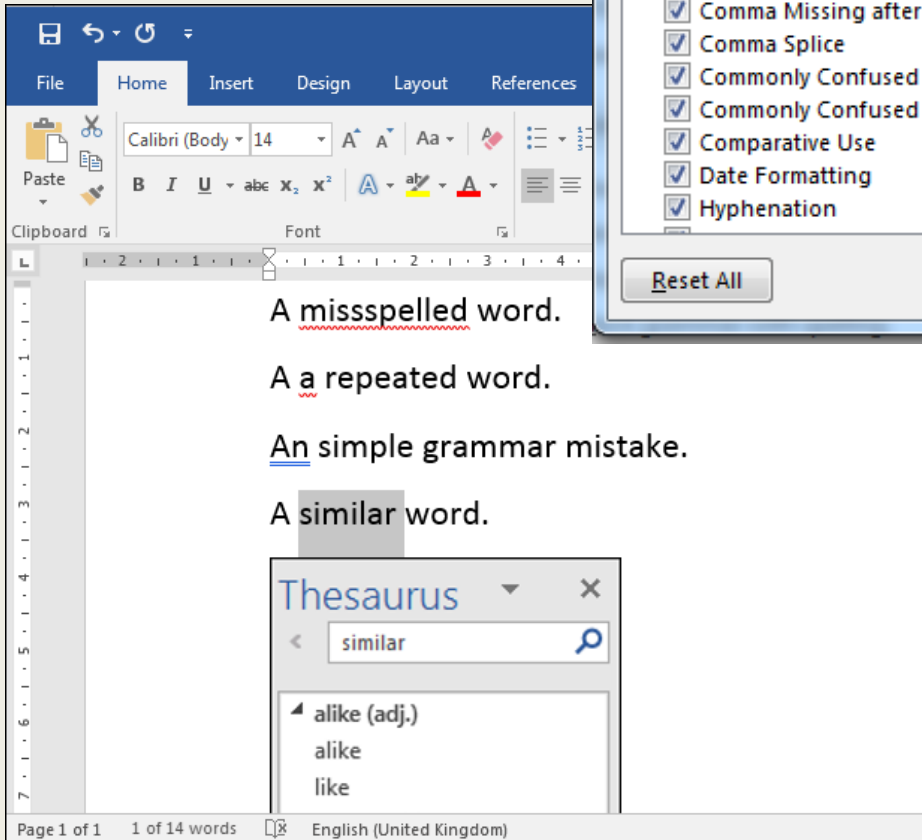
Limitations

- Corpora directly: require skills most don't have
- If curated (mostly dictionaries): **limited coverage**, examples
- Information may be **hidden away** within the entry
- If not curated (SkELL, flax, netspeak.org, other corpus-based): **noise**
- Not specific to Academic English (exceptions: LEAD, OLDAE)
- Users have to interrupt their writing to consult the collocation resource

WRITING ASSISTANTS



MS Word



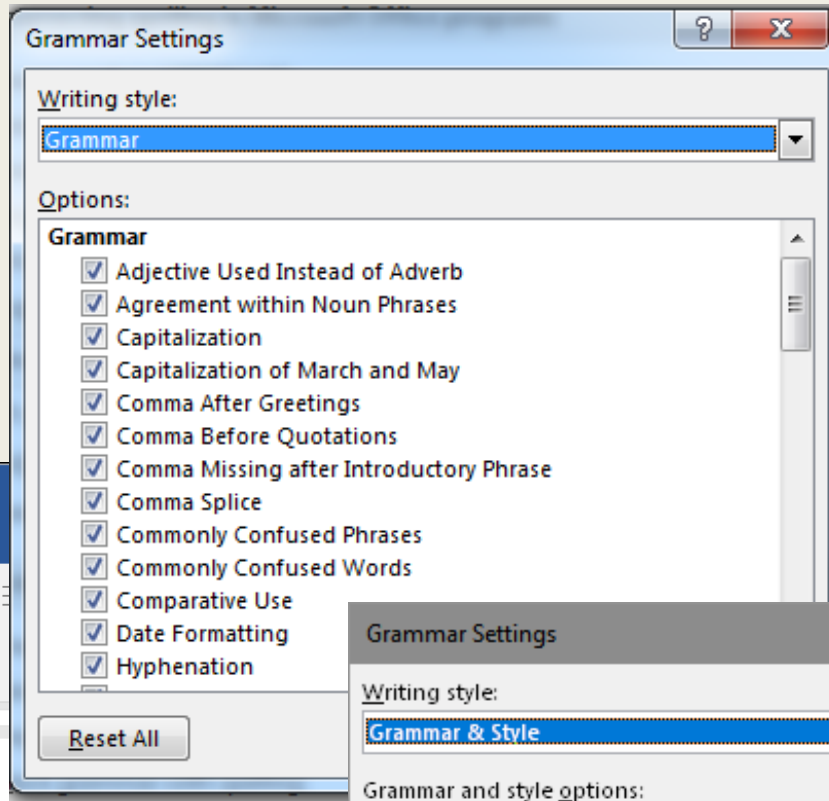
The screenshot shows the MS Word interface with the Home ribbon selected. The font is Calibri (Body) 14. The document content includes:

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

A Thesaurus window is open, showing the search term "similar" and the following suggestions:

- alike (adj.)
- alike
- like

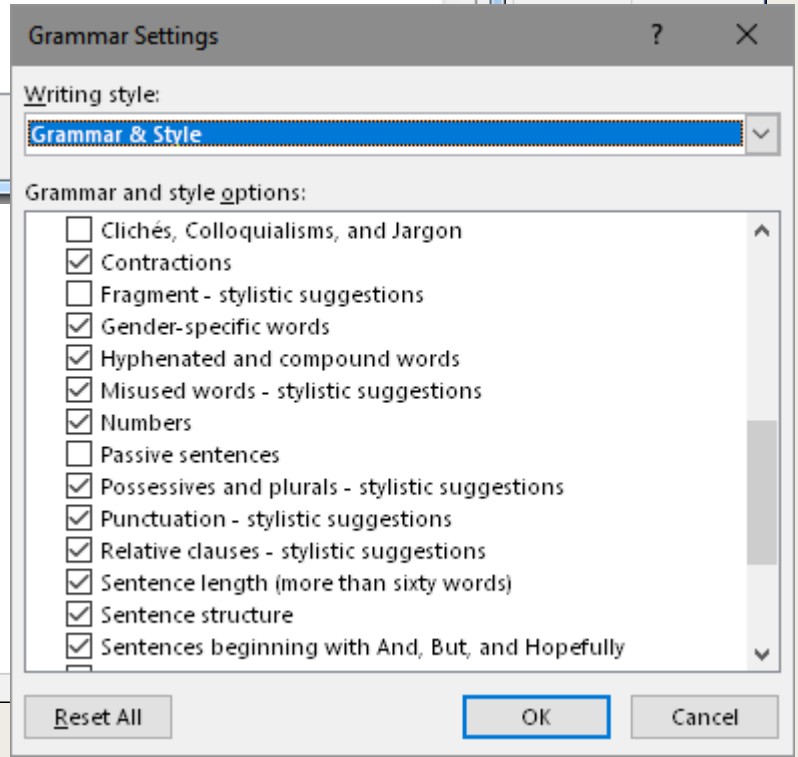
Page 1 of 1 1 of 14 words English (United Kingdom)



The screenshot shows the Grammar Settings dialog box with the "Writing style" set to "Grammar". The "Options" section is expanded to "Grammar", and the following options are checked:

- 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

Reset All

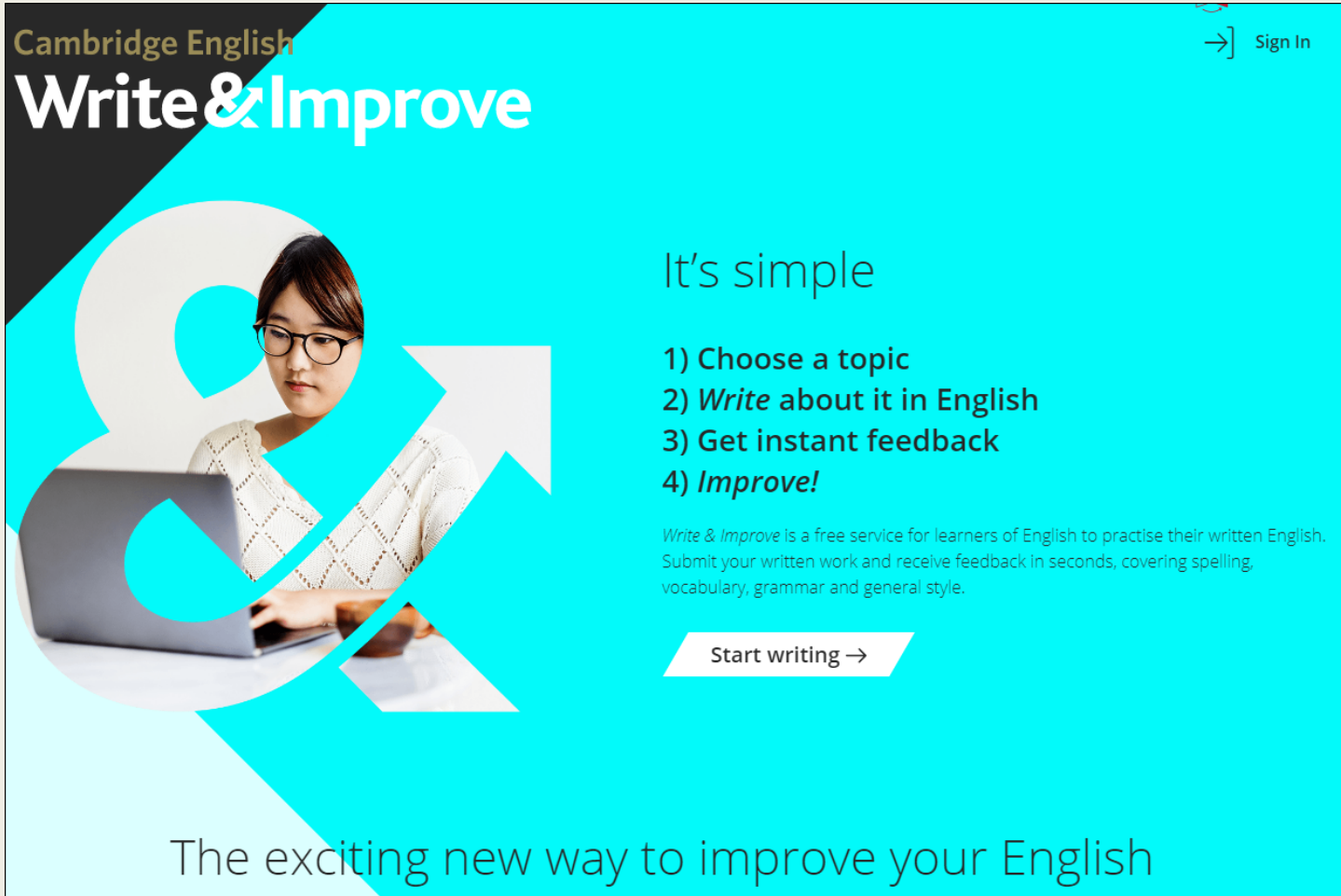


The screenshot shows the Grammar Settings dialog box with the "Writing style" set to "Grammar & Style". The "Grammar and style options" section is expanded, and the following options are checked:

- Contractions
- Gender-specific words
- Hyphenated and compound words
- Misused words - stylistic suggestions
- Numbers
- Possessives and plurals - stylistic suggestions
- Punctuation - stylistic suggestions
- Relative clauses - stylistic suggestions
- Sentence length (more than sixty words)
- Sentence structure
- Sentences beginning with And, But, and Hopefully


Reset All OK Cancel

Cambridge English Write & Improve



Cambridge English
Write & Improve

→] Sign In



It's simple

- 1) Choose a topic
- 2) *Write* about it in English
- 3) Get instant feedback
- 4) *Improve!*

Write & Improve is a free service for learners of English to practise their written English. Submit your written work and receive feedback in seconds, covering spelling, vocabulary, grammar and general style.

Start writing →

The exciting new way to improve your English

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

GENERAL

ACADEMIC

OVERUSE

LEARNER

WriteAway

In the **present** contribution, we explain the patterns

less

more patterns

less

more examples

write

edit

English

英漢

英和

[V] present something in → present something 48

*assembly [presented the constitution in] 24 , 1979 , =>assembly [presented the constitution] October 24 , 1979 ,
 , [presented a Norwegian lumberjack in] Upper Midwest=> , [presented a Norwegian lumberjack] the Upper
Midwest*

[V] present something of → present something 28

*able to [present proof of] the presumed licensee has=>able to [present proof] the presumed licensee has
Doyle [presents a comparison of] debuting character=>Doyle [presents a comparison] his debuting character*

[V] present with something → present something 26

*and [presents with a significantly worse deformity] postural kyphosis=>and [presents a significantly worse
deformity] postural kyphosis
often [present with an internalized , inattentive sub-type] the disease which=>often [present an internalized ,
inattentive sub-type] the disease which*

Existing writing assistants: problems

- Mostly *reactive* feedback (not real-time)
- Often corpus-based using NLP: noise
- Not specific to Academic English
- Not about collocation
- Usability issues: intrusive, unintuitive



OUR SOLUTION

Project led by University of Surrey (A. Frankenberg-Garcia, G. Rees)
with Bangor University (J. Roberts, N. Sharma)
Adam Mickiewicz University in Poznań (R. Lew)
Funded by Arts & Humanities Research Council UK (AH/P003508/1)

An alternative solution



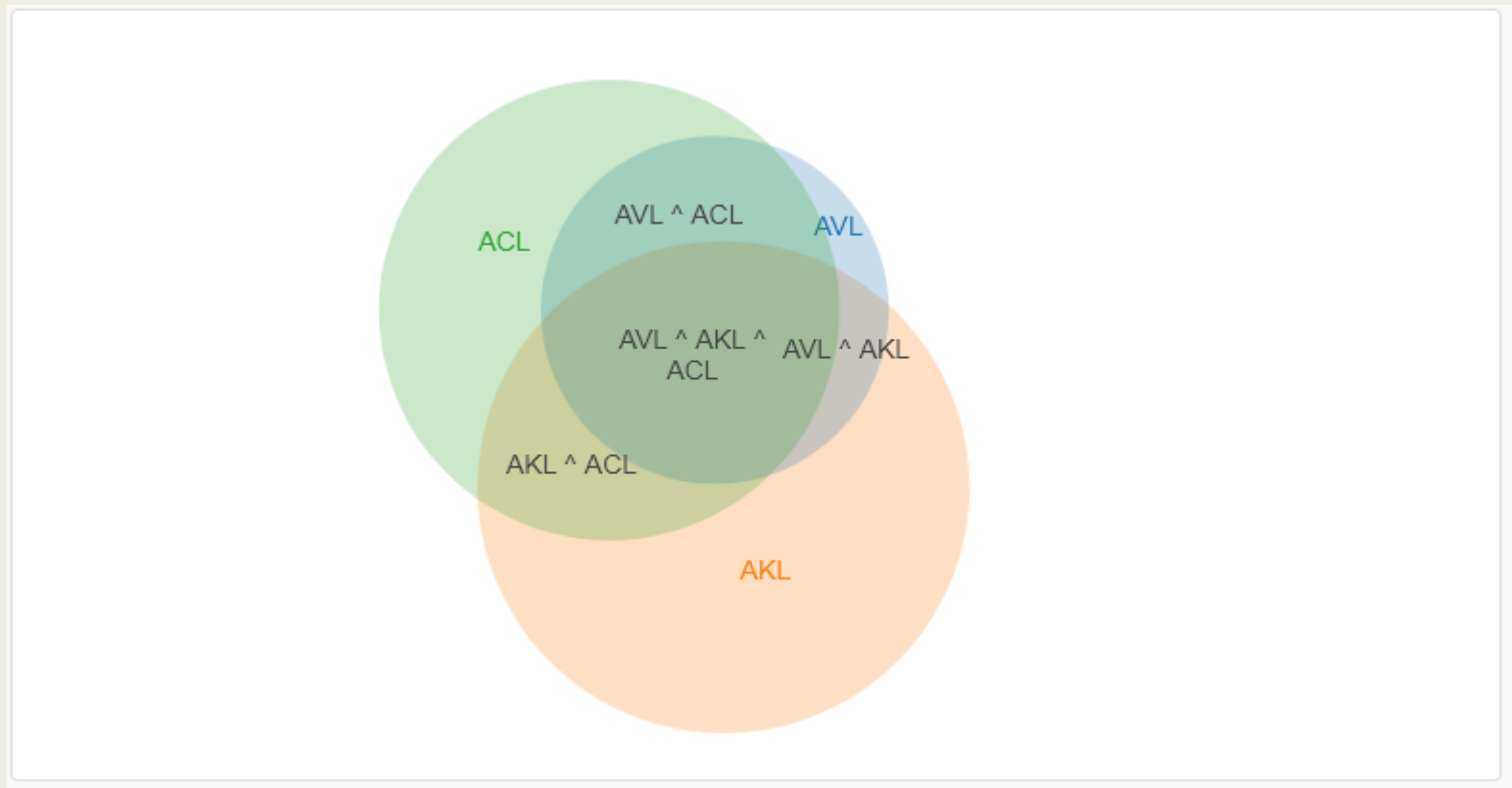
COLLOCAID

- An intuitive collocation-aware lexicographic resource integrated with digital writing environments that:
 - *requires little training or instruction*
 - *raises language awareness*
- Focus on Academic English (EAP) collocations:
 - *carry out + research*
 - *statistically + significant*
 - *further + studies*

Our set of collocate bases (nodes)

- Based on three recent research-based sets of EAP vocabulary
 - *Academic Vocabulary List (AVL, Gardner & Davies 2014, based on COCA); narrowed down as per Durrant (2016), who found that just 427 AVL items were frequent in a broad range of disciplines in university student writing (BAWE corpus)*
 - *Academic Keyword List (AKL, Paquot 2010), based on the LOCNESS Corpus (Granger 1998) of British and American written assignments*
 - *Academic Collocation List (ACL, Ackermann & Chen 2013) found in the appendix to the Longman Collocations Dictionary*
- Criterion for inclusion: item present in more than one of these 3 lists

At least two out of three lists: 512 nodes



512 nodes selected

Noun bases
282

Adjectives &
modifiers:

e.g. a complex
system,
a classification
system,
an existing system,
etc.

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

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

Prepositions:
e.g. within a system,
a system for something

Verb bases
136

Adverbs:

e.g. undoubtedly
contribute,
contribute significantly,
contribute substantially,
etc.

Prepositions:
e.g. contribute
to

Adjective bases
94

Adverbs:

e.g. statistically
significant,
particularly significant,
etc.

Oxford Corpus of Academic English in the Sketch Engine

Sketch Engine [Oxford Corpus of Academic English \(April 2012\)](#)

- Home
- Search
- Word list
- Word sketch
- Thesaurus
- Sketch diff
- Corpus info
- My jobs
- User guide

- Save
- Change options
- Cluster
- Sort by freq
- Hide gramrels
- More data
- Less data
- Sketch grammar
- Translate
 - Arabic
 - Bulgarian
 - Czech
 - Dutch
 - French
 - German
 - Italian
 - Polish
 - Russian
 - Spanish
- Menu position

important *(adjective)*
Oxford Corpus of Academic English (April 2012) freq = [55,647](#) (659.13 per million)

ADV ADJ*		12.29
particularly +	912	10.94
is particularly important		
equally +	401	10.40
. equally important		
especially +	430	10.35
is especially important		
increasingly +	398	10.17
an increasingly important		
very +	1,495	10.03
very important		
extremely +	296	9.75
is extremely important		
as +	839	9.52
as important as		
so +	361	9.31
is so important		
critically +	124	9.13
critically important		
vitality +	118	9.11
is vitality important		
potentially +	138	8.79
potentially important		
crucially	78	8.52
is crucially important		
really	75	8.19

Adding collocates and examples

- logDice score of ≥ 5
- minimum co-occurrence frequencies
 - *10 for lexical collocates*
 - *100 for prepositions*
- remove collocates that are too general (e.g. *own* and *good*)
- remove collocates restricted to a small number of disciplines (e.g. *entrepreneurial ability* restricted to business studies)
- select three examples per each collocate-base pair
- replace proper nouns with pronouns, change date, numerical and name references
- editing done in a cloud-based spreadsheet

Excerpt from the database

BASE	POS	RELATION	COLLOCATE	CO-	ASSOCI	EXAMPLE1
equal	j	ADV ADJ*	roughly	108	11.42	the latter two groups had roughly equal rates of break
equal	j	ADV ADJ*	exactly	62	10.85	total costs and our total revenues are exactly equal
equal	j	ADV ADJ*	nearly	57	9.89	three experiments were performed using nearly equal
equal	j	ADV ADJ*	almost	69	9.26	men and women are apparently almost equal now in t
equal	j	ADV ADJ*	formally	10	8.35	this occurs where treatment is formally equal
equal	j	ADV ADJ*	necessarily	14	7.81	attachment does not necessarily equal ownership
equal	j	ADV ADJ*	relatively	20	5.48	all household members have relatively equal access to
important	j	ADV ADJ*	particularly	912	10.94	the sensitivity of a test is particularly important
important	j	ADV ADJ*	equally	401	10.4	the two stages are equally important and interlinked
important	j	ADV ADJ*	especially	430	10.35	especially important were the localization of brain and
important	j	ADV ADJ*	increasingly	398	10.17	public relations is becoming an increasingly important
important	j	ADV ADJ*	very	1 495	10.03	it is very important for an economy to be stable
important	j	ADV ADJ*	extremely	296	9.75	they see work and its consequences as extremely imp
important	j	ADV ADJ*	critically	124	9.13	determine which points of critically important informa
important	j	ADV ADJ*	vitaly	118	9.11	the link between the two concepts is therefore vitaly

Example scenarios



COLLOCAID

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

A small increase of temperature

✗ Do not say **increase of** something
✓ Say **increase in** something

Example scenarios



- Not just correcting reactively, but guiding the process

Research done as a result of...

- conduct research ⊕
- research suggests
- emp
- rese
- 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

Example scenarios



- Customizable
 - Option to switch off real-time help and check text at the end
 - We plan to run usability tests on a prototype
- Research done as a result of...

- conduct research +
- research suggests +
- empirical research +
- research on something +

Hide

Restore hidden prompts

- research
- hypothesis
- analyse

First prototype implementation

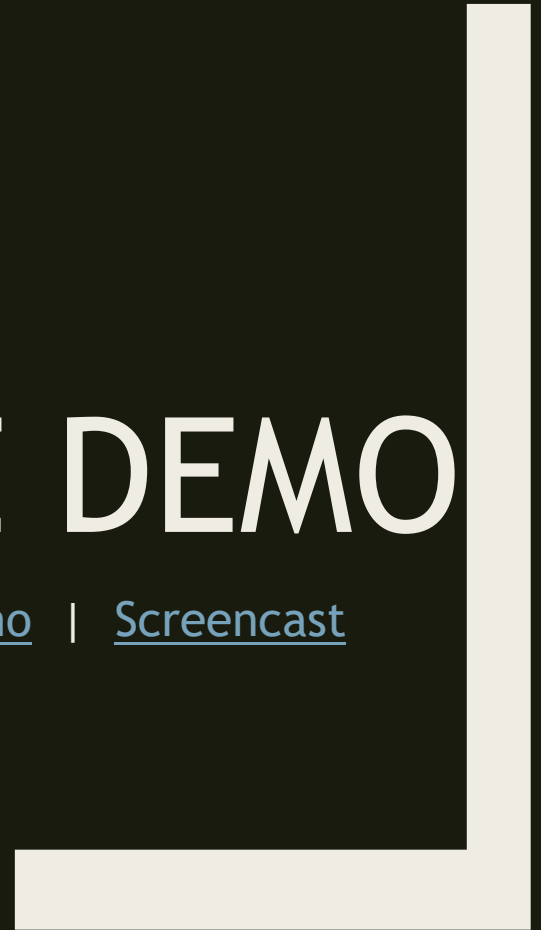


COLLOCAid

- Custom plugin for the TinyMCE text editor
 - *TinyMCE is*
 - open-source
 - lightweight
 - extensible
 - written in JavaScript
 - intended for web applications
 - client-side (web browser does the work)

PROTOTYPE DEMO

[Local File](#) | [On Vimeo](#) | [Live Demo](#) | [Screencast](#)



How will ColloCaid be different?

- Lexicographic data will be curated so that
 - *we provide just the right amount (less is more)*
- Specific focus on Academic English
- When writing, users won't get distracted with:
 - *irrelevant or misleading information*
 - *information overload*
- Prompts will be discreet and customizable, so that
 - *they don't fill too much screen space*
 - *users don't have to deal with lengthy explanations*
- Metalanguage will be avoided, so users don't have to struggle with grammatical terminology

The ColloCaid team

(in the middle of a Skype conference)



UNIVERSITY OF
SURREY



PRIFYSGOL
BANGOR
UNIVERSITY