Combining learner needs, lexicographic data and digital writing environments

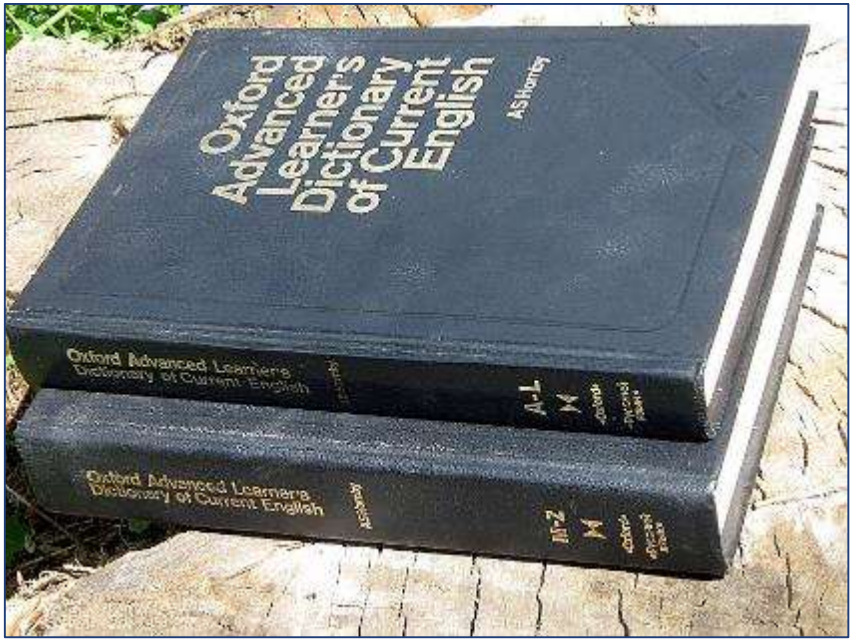
Ana Frankenberg-Garcia
a.frankenberg-garcia@surrey.ac.uk

Dictionaries in the past

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

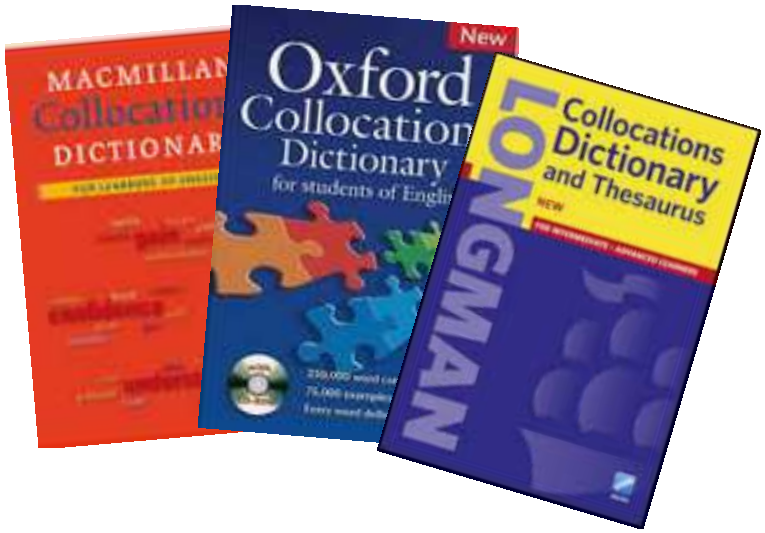
The user paradigm

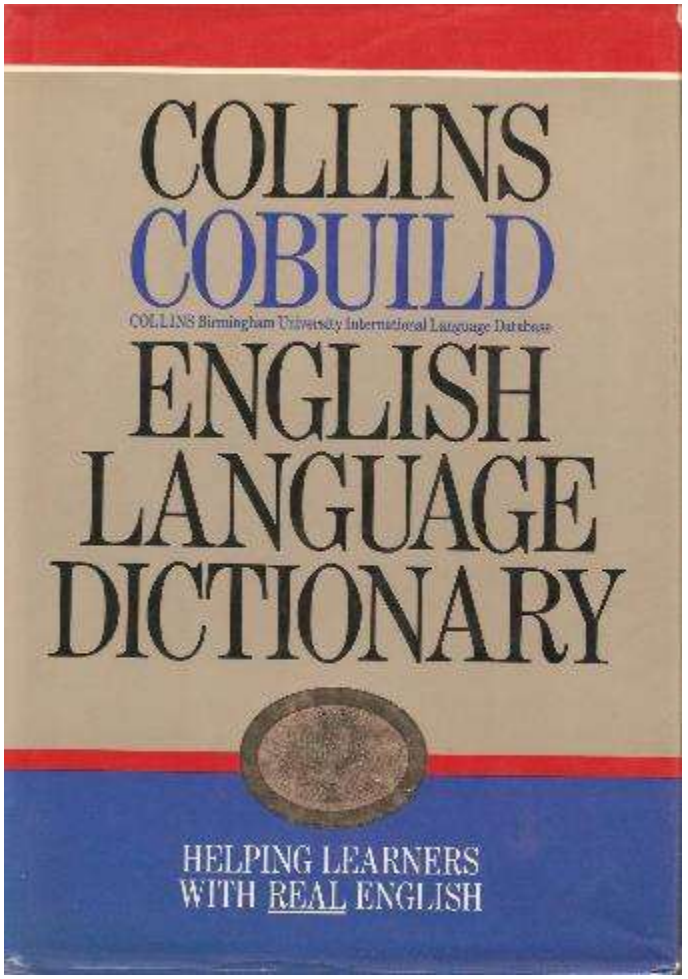
- Dictionaries for learners
- Headword list selection
- Defining vocabulary
- Usage labels
- Usage patterns
- Pronunciation



The corpus paradigm

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

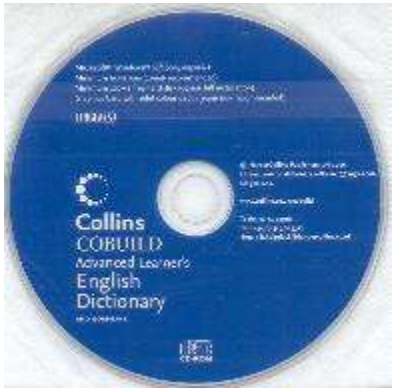


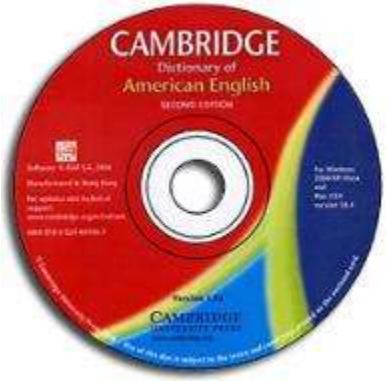


- New types of dictionaries

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 real-time paradigm

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

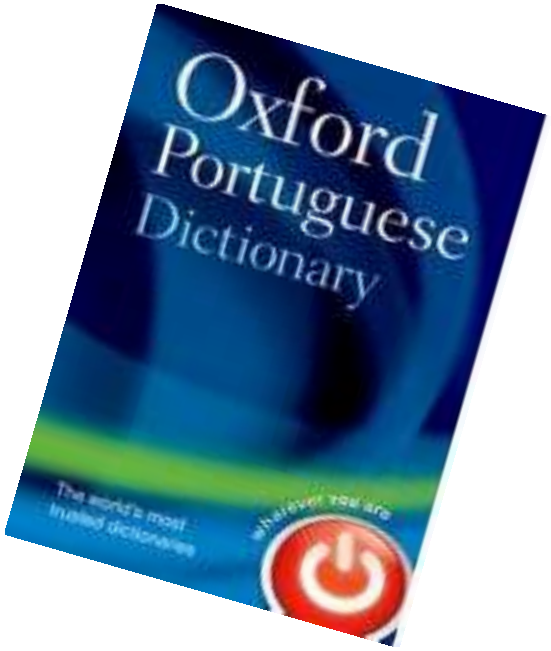
And yet…

- Spectacular developments have not had a dramatic impact on actual dictionary-user behaviour.
- Dictionaries used mainly just for meanings
- Both paper-based and digital dictionaries remain by and large underused

In second language and translation, marked preference for bilingual look-ups

(e.g. Frankenberg-Garcia 2005)

- L2-L1 equivalents for comprehension
- L1-L2 equivalents for production



Language users in general are not fully aware of

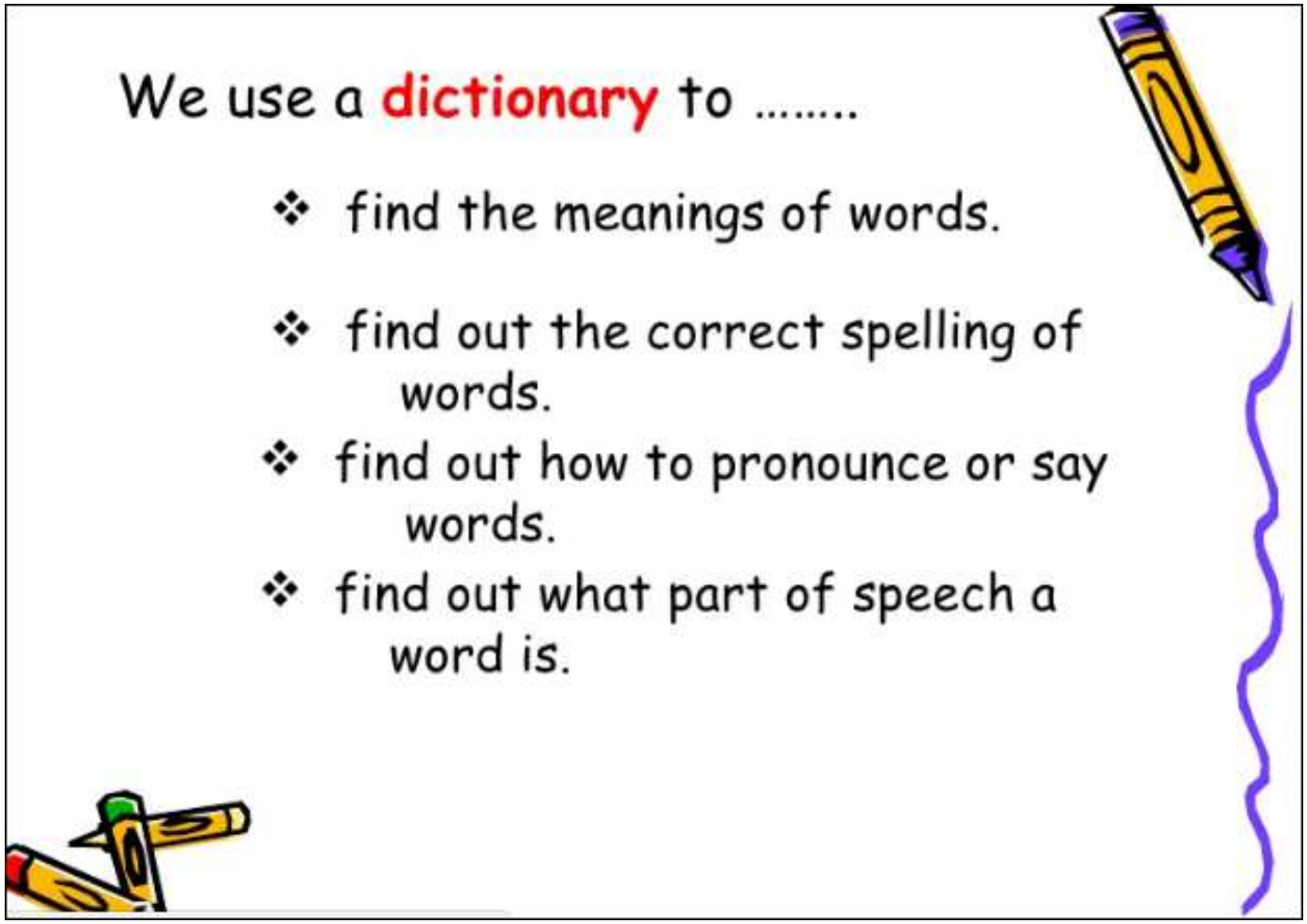
- What else current dictionaries can offer them
  - e.g. most users didn't know they could use dictionaries look up collocations (Frankenberg-Garcia 2011)

- 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)

Addressing dictionary underuse

- 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)

Teaching dictionary use



https://www.slideshare.net/acw1217/3rd-dictionary-guide-words [31/05/2017]

Teaching dictionary use

https://www.wikihow.com/Use-a-Dictionary [6 /11/2017]

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…

An alternative solution

Dictionary-like tools that

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

- www.collocaid.uk
- Project led by Surrey (Frankenberg-Garcia), with Bangor (Roberts), Poznan (Lew)
- Funded by AHRC AH/P003508/1

What is Collocaid?

- An intuitive lexicographic resource integrated with digital writing environments to help learners write more idiomatically
- Focus on Academic English (EAP) collocations
  - The way words combine in academic English
    - *carry out + research*
    - *statistically + significant*
    - *further + studies*

Why collocations?

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)

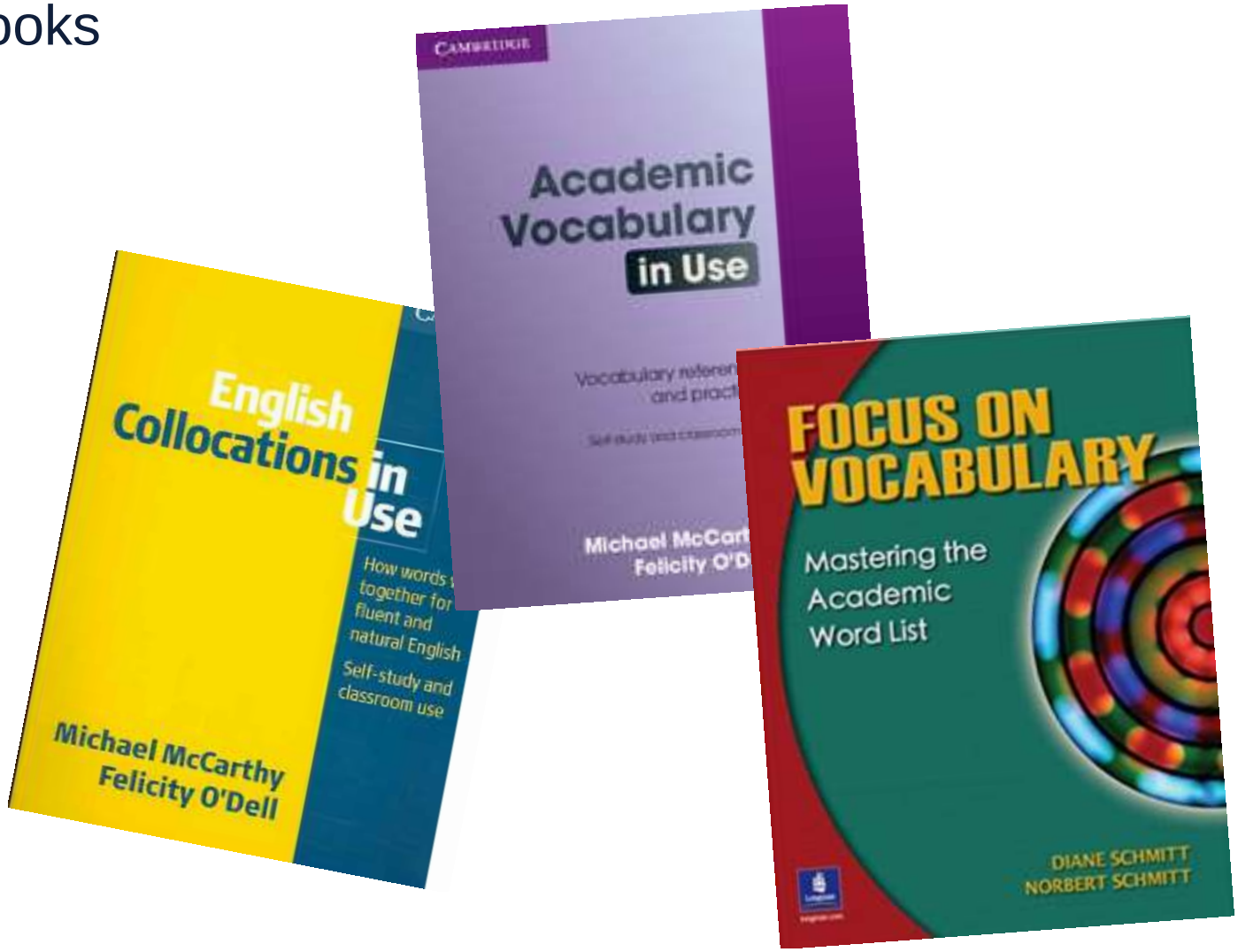
Why Academic English?

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 the dissemination of science (Jenkins 2014)
4. There are no native speakers of academic language (Hyland 2006, Kosem 2010, Frankenberg-Garcia 2017)

Of use to both native and non-native writers who do not have sufficient command of EAP collocations

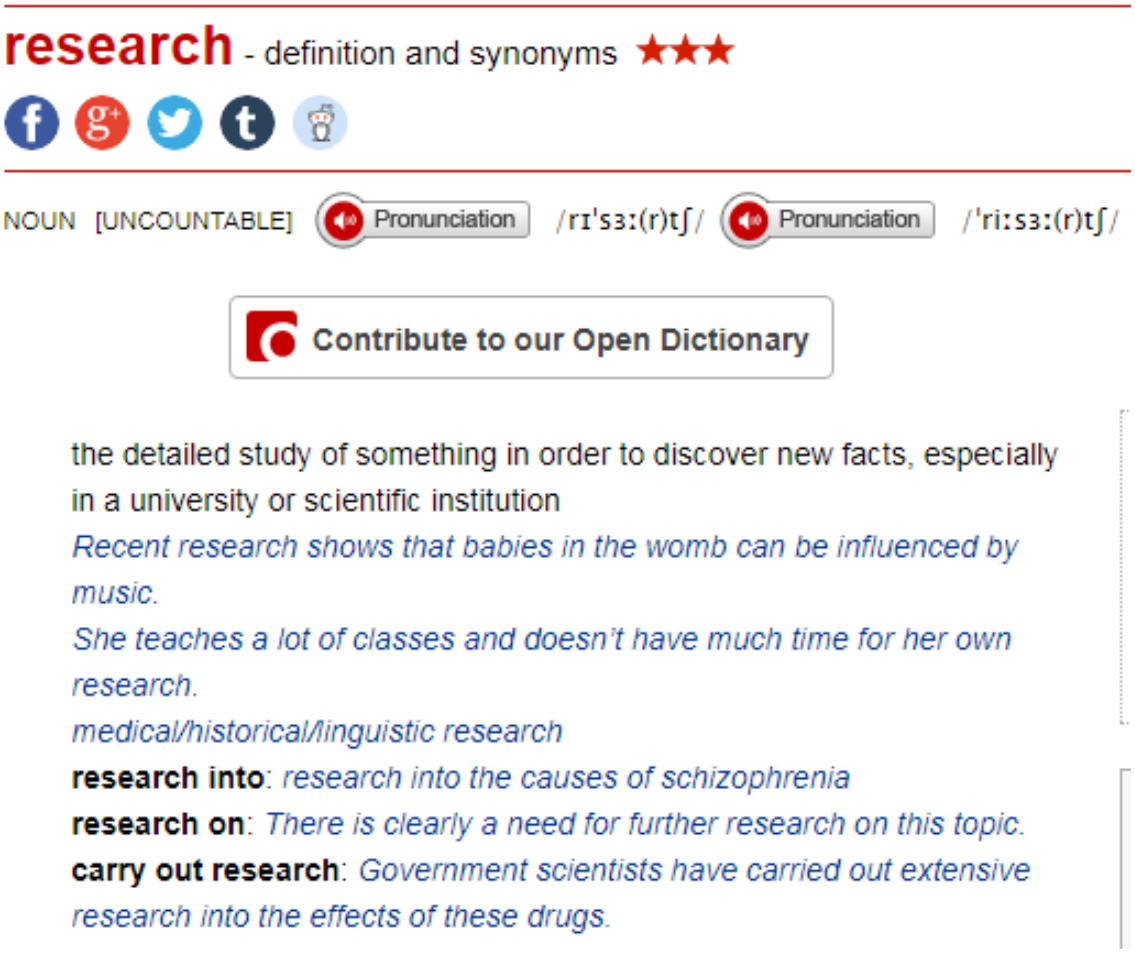
UNIVERSITY OF SURREY

Textbooks



General English Dictionaries for Learners

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



research - definition and synonyms ★★★

NOUN [UNCOUNTABLE] Pronunciation /rɪ'sɜ:(r)tʃ/ Pronunciation /'ri:sɜ:(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.

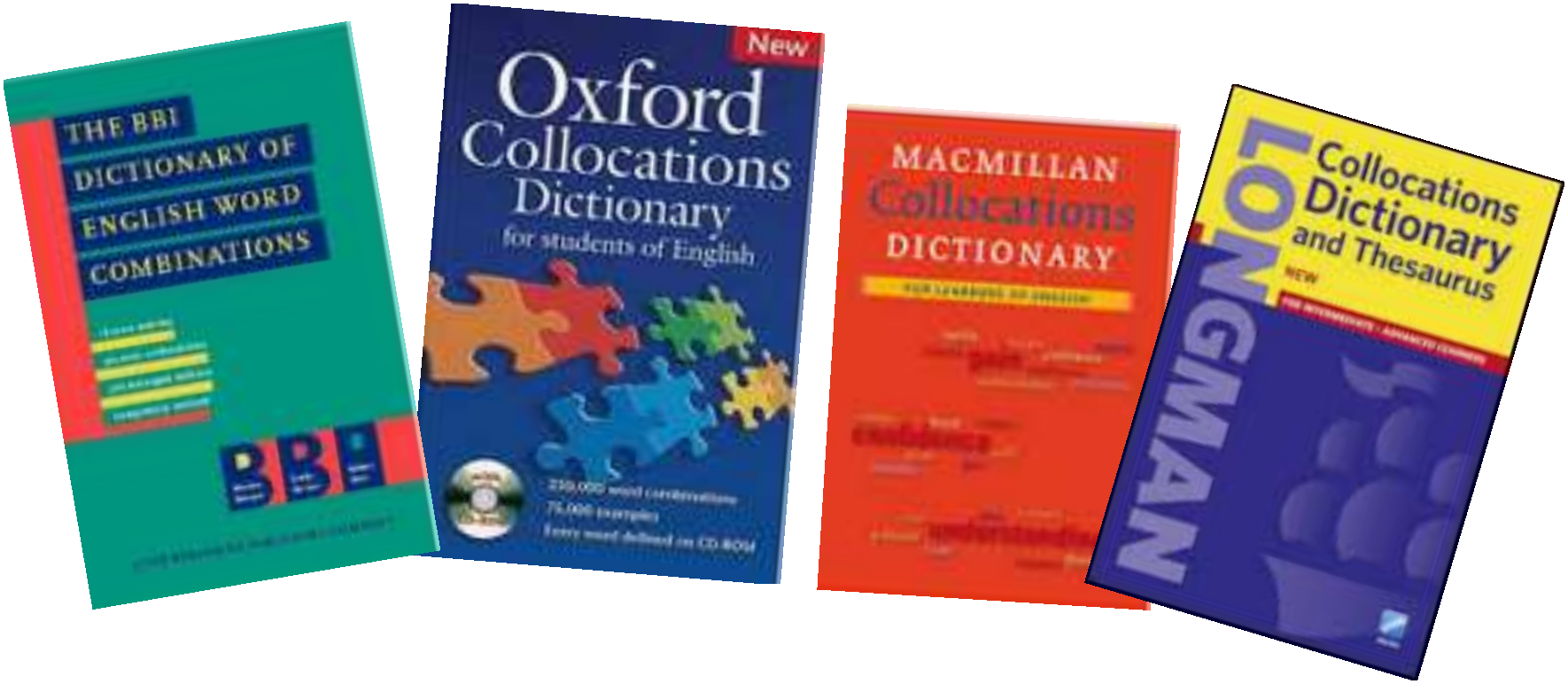
medical/historical/linguistic research

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.*

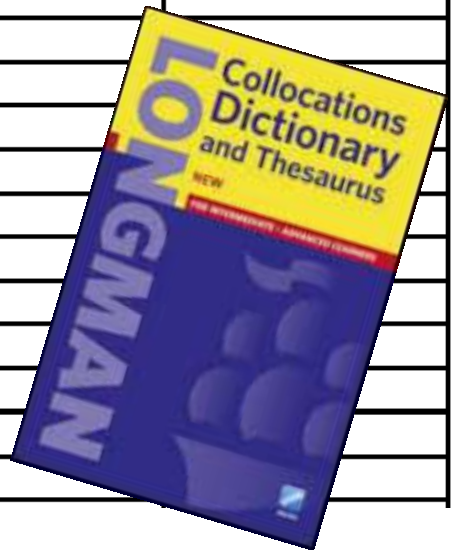
General English Collocation Dictionaries



Existing help with EAP collocations

The Academic Collocation List

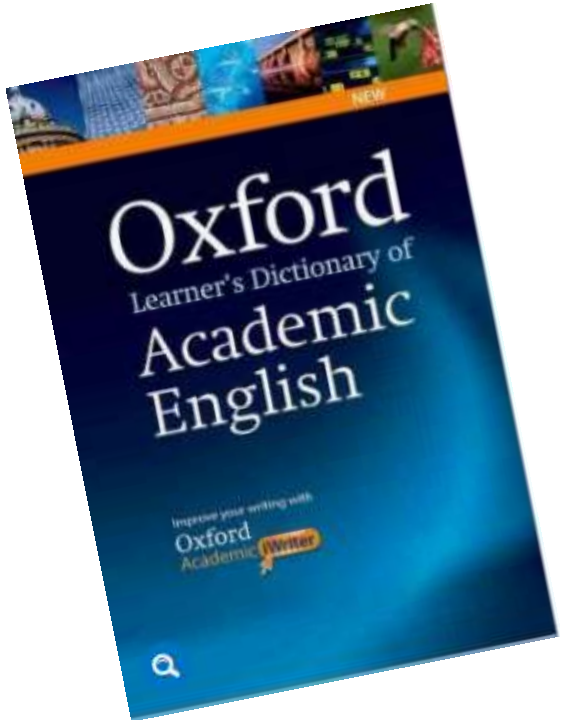
#	Addition	Component I	POS I	Component II	POS II	Addition
1		abstract	adj	concept	n	
2		academic	adj	achievement	n	
3		academic	adj	career	n	
4	(in)	academic	adj	circles	n	
5		academic	adj	community	n	
6		academic	adj	debate	n	
7		academic	adj	discipline	n	
8		academic	adj	discourse	n	
9		academic	adj	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)

Existing help with EAP collocations

- Based on the Oxford Corpus of Academic English
- “700 of the most important, collocationally prolific words [given] a separate section of the entry… in the style of a collocations dictionary” (Lea 2014:185)
- Example sentences derived from the corpus

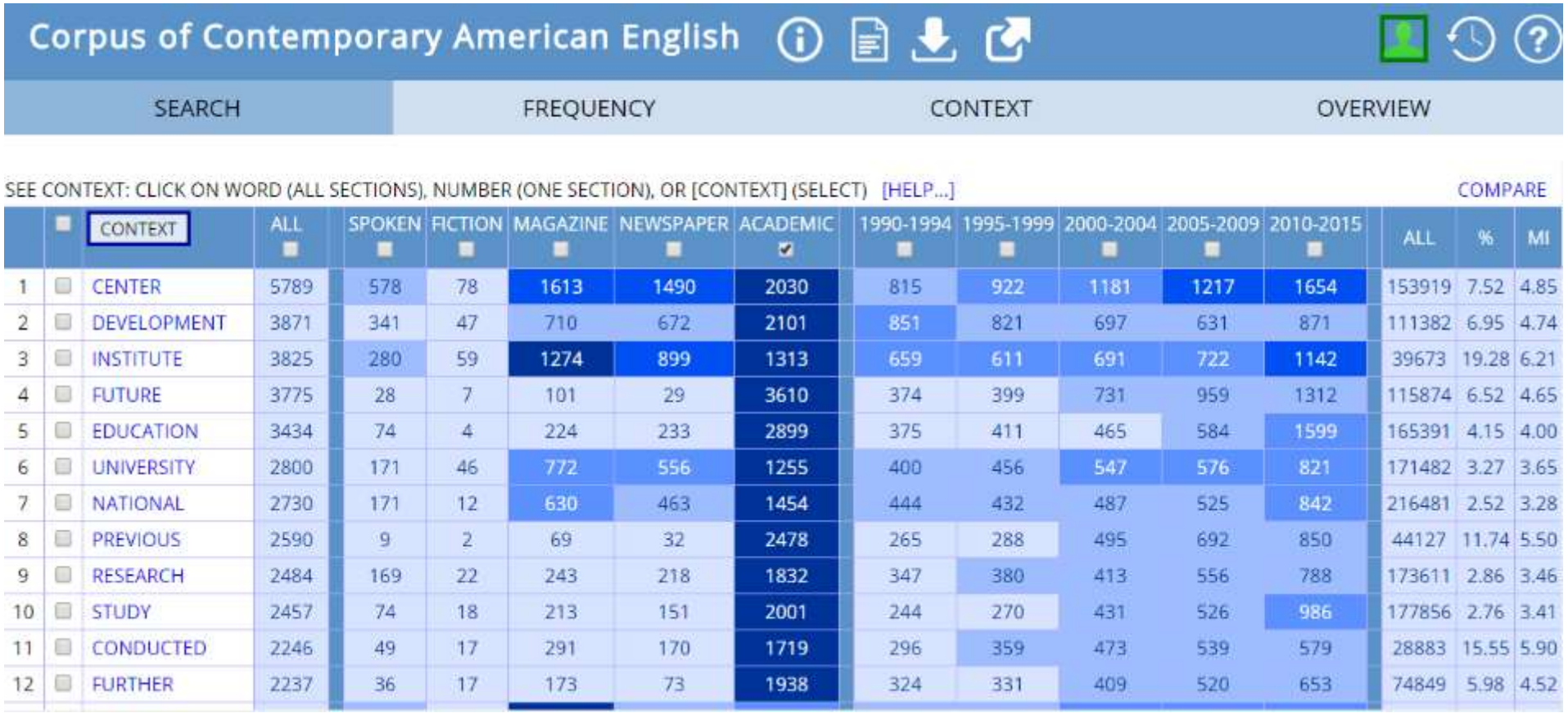


Academic Formulas List

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

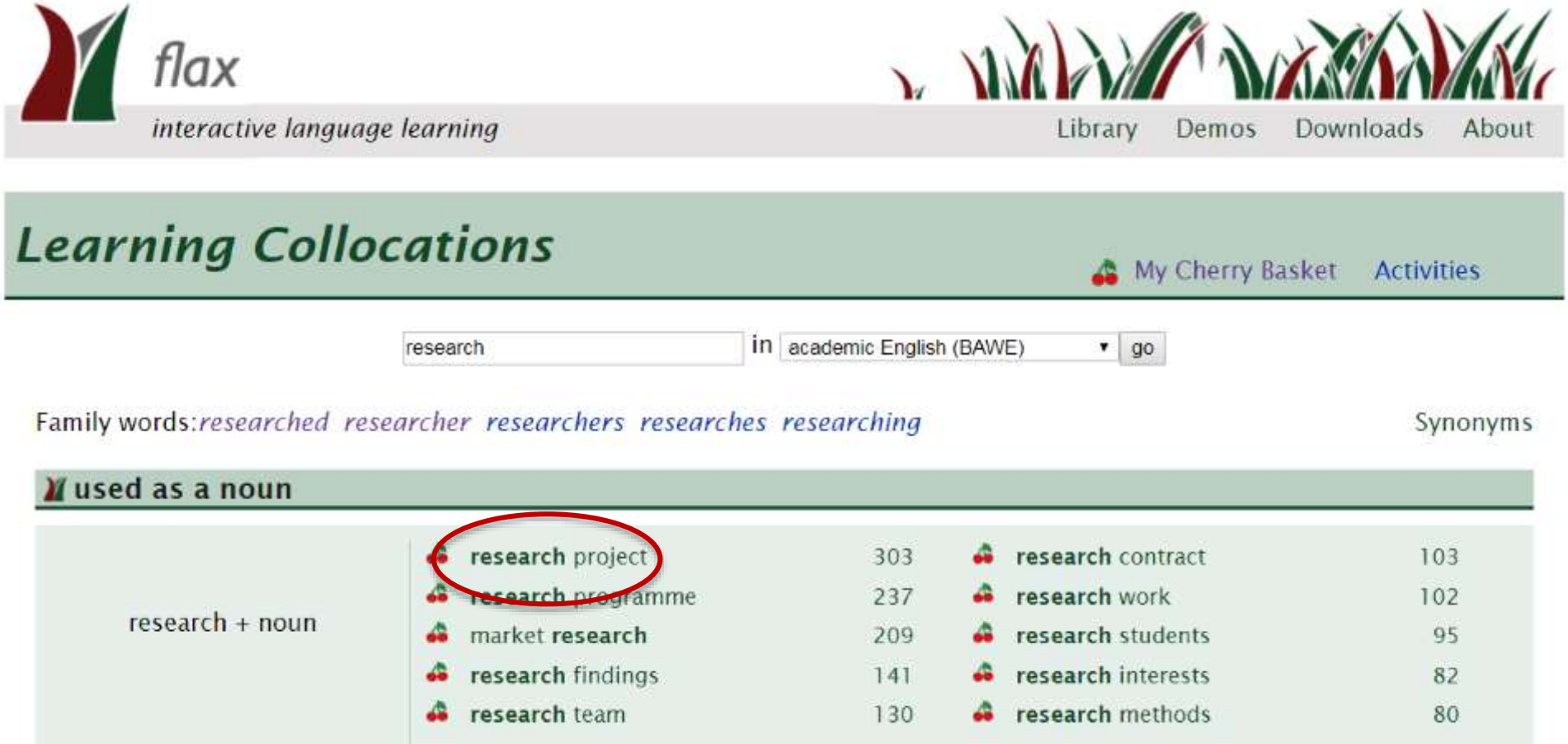
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...] COMPARE

	CONTEXT	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	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
12	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)

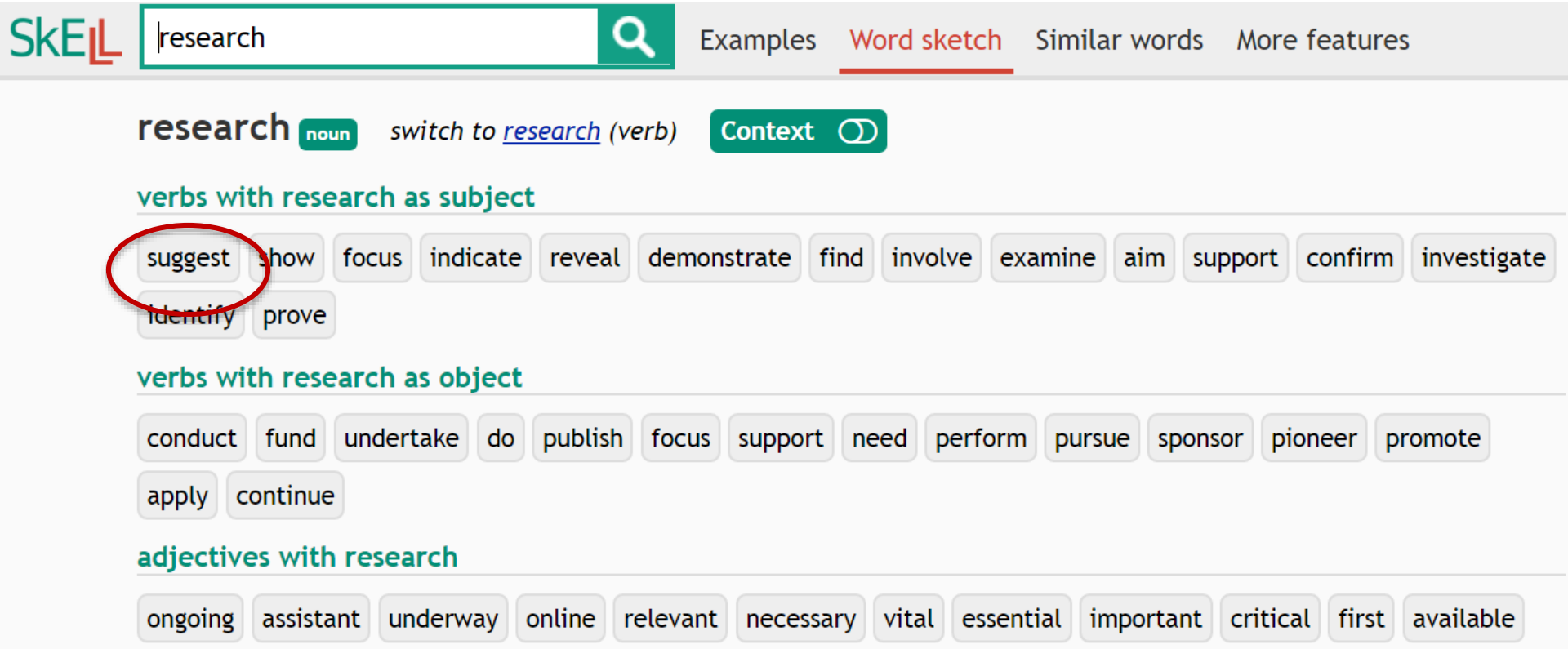
Existing help with EAP collocations



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

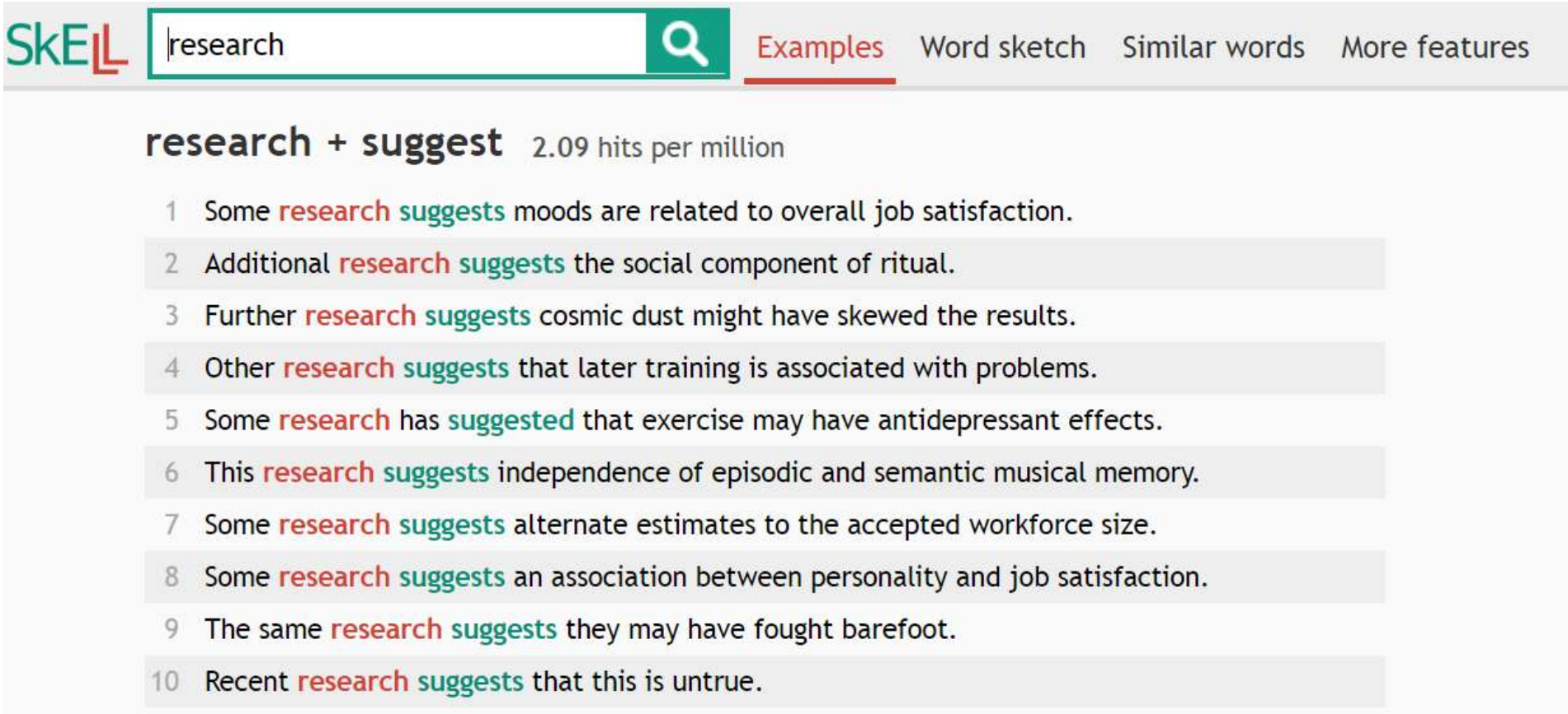
Existing help with EAP collocations

Existing help with EAP collocations



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

Existing help with EAP collocations



Existing help with EAP Collocations

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

Existing help with EAP collocations

In summary

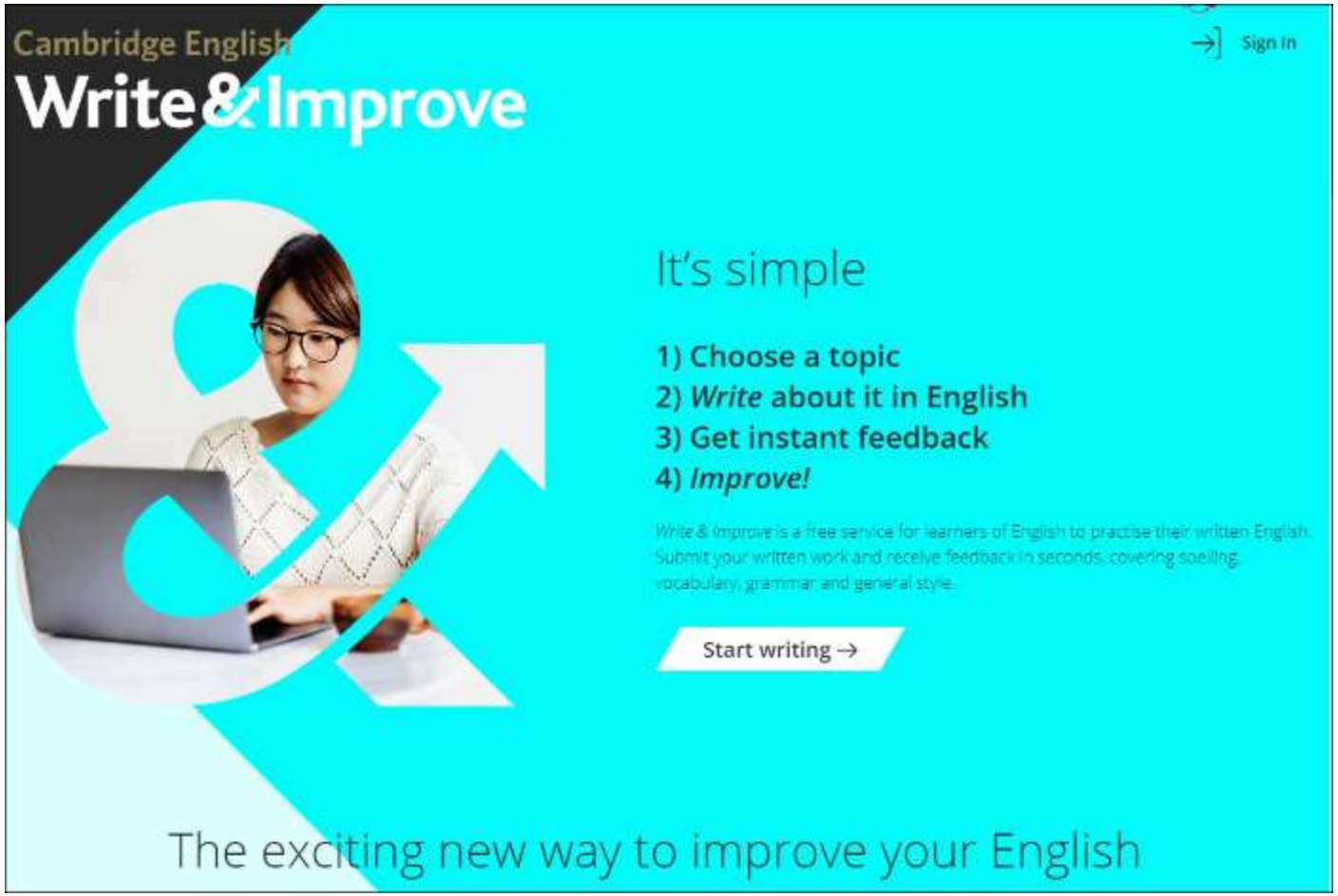
- Excellent general English resources (which include EAP)
- Excellent specific EAP resources

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)
- 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)

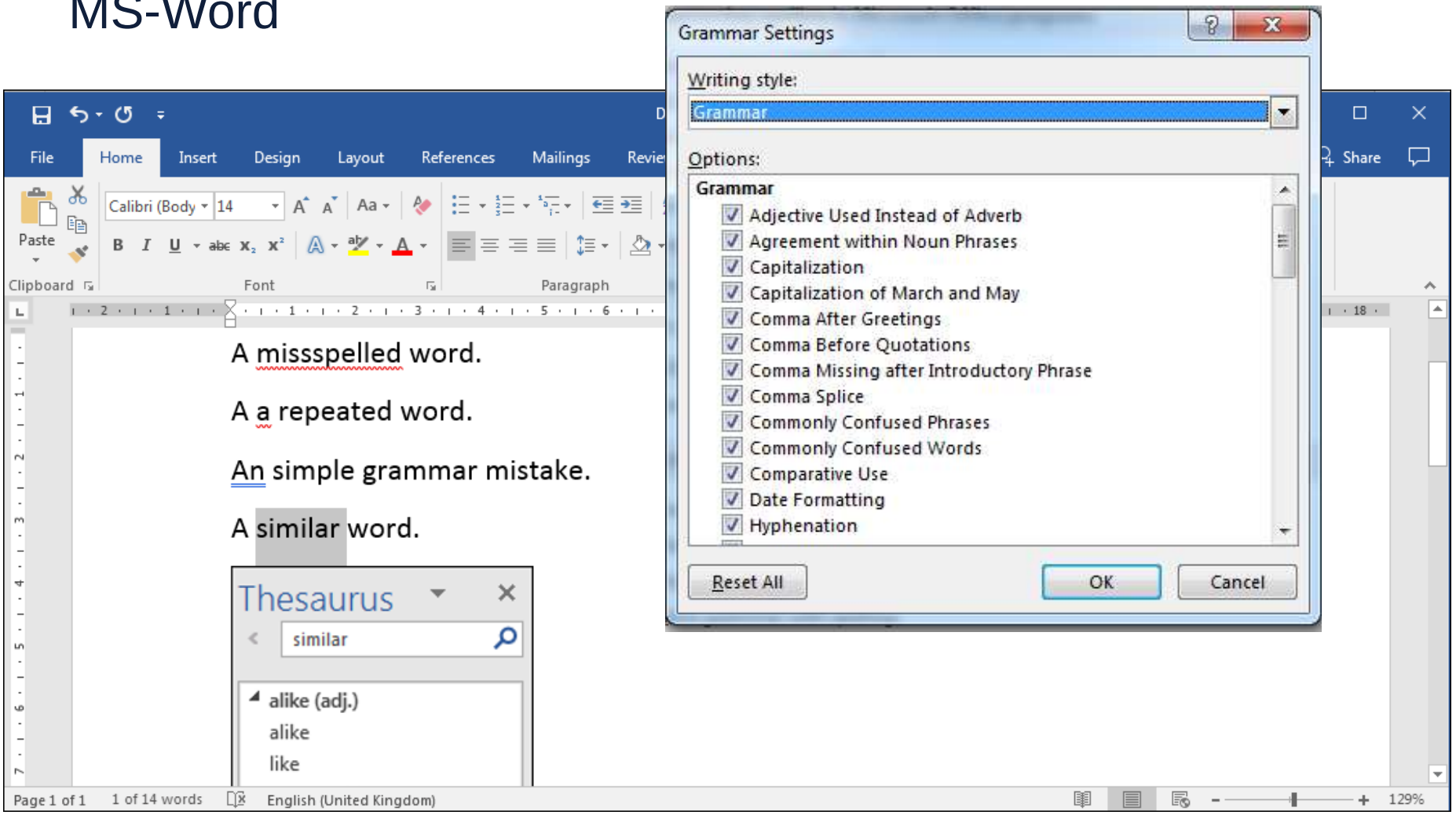
Existing real-time help with writing

Write & Improve https://writeandimprove.com/



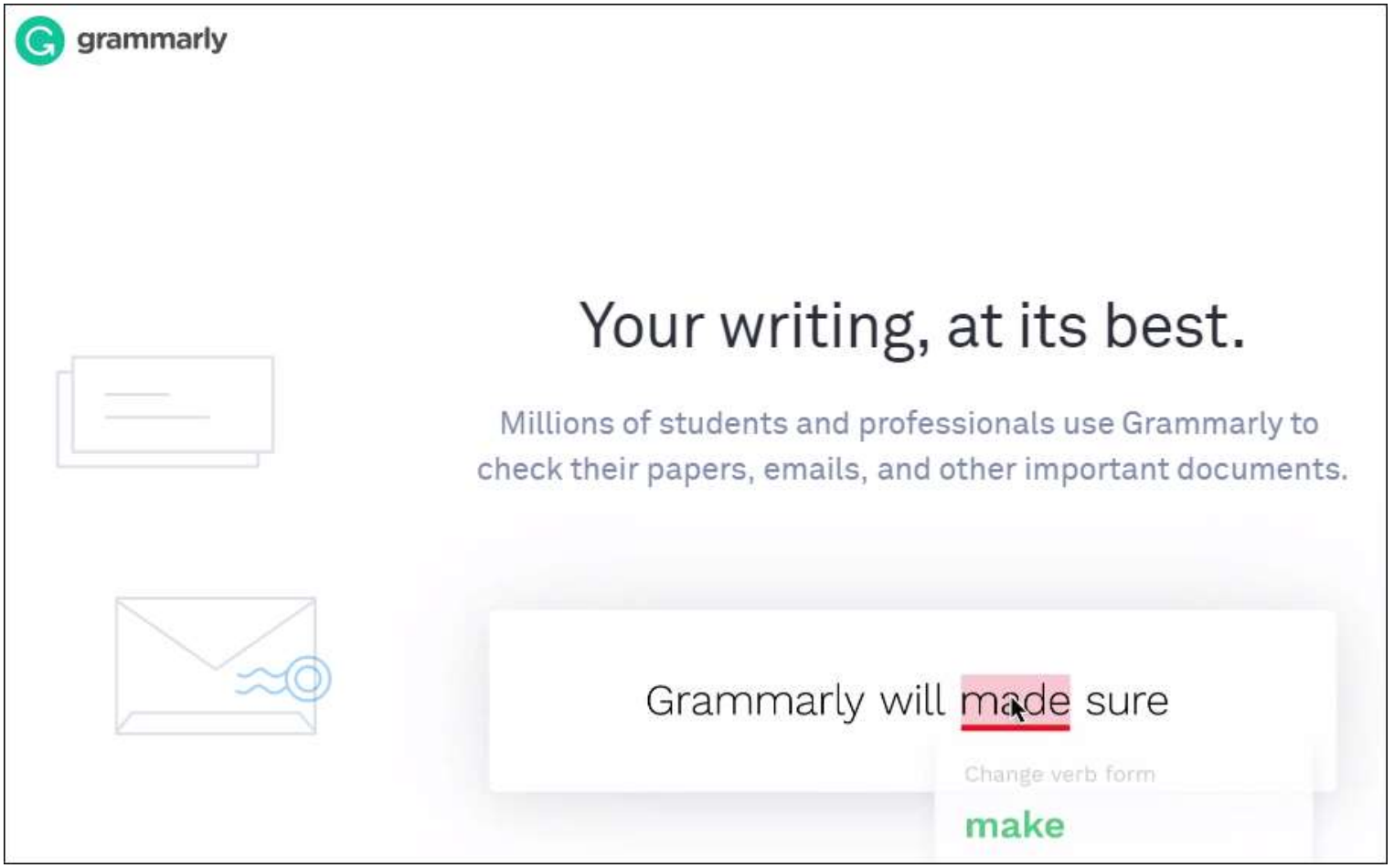
Existing real-time help with writing

MS-Word



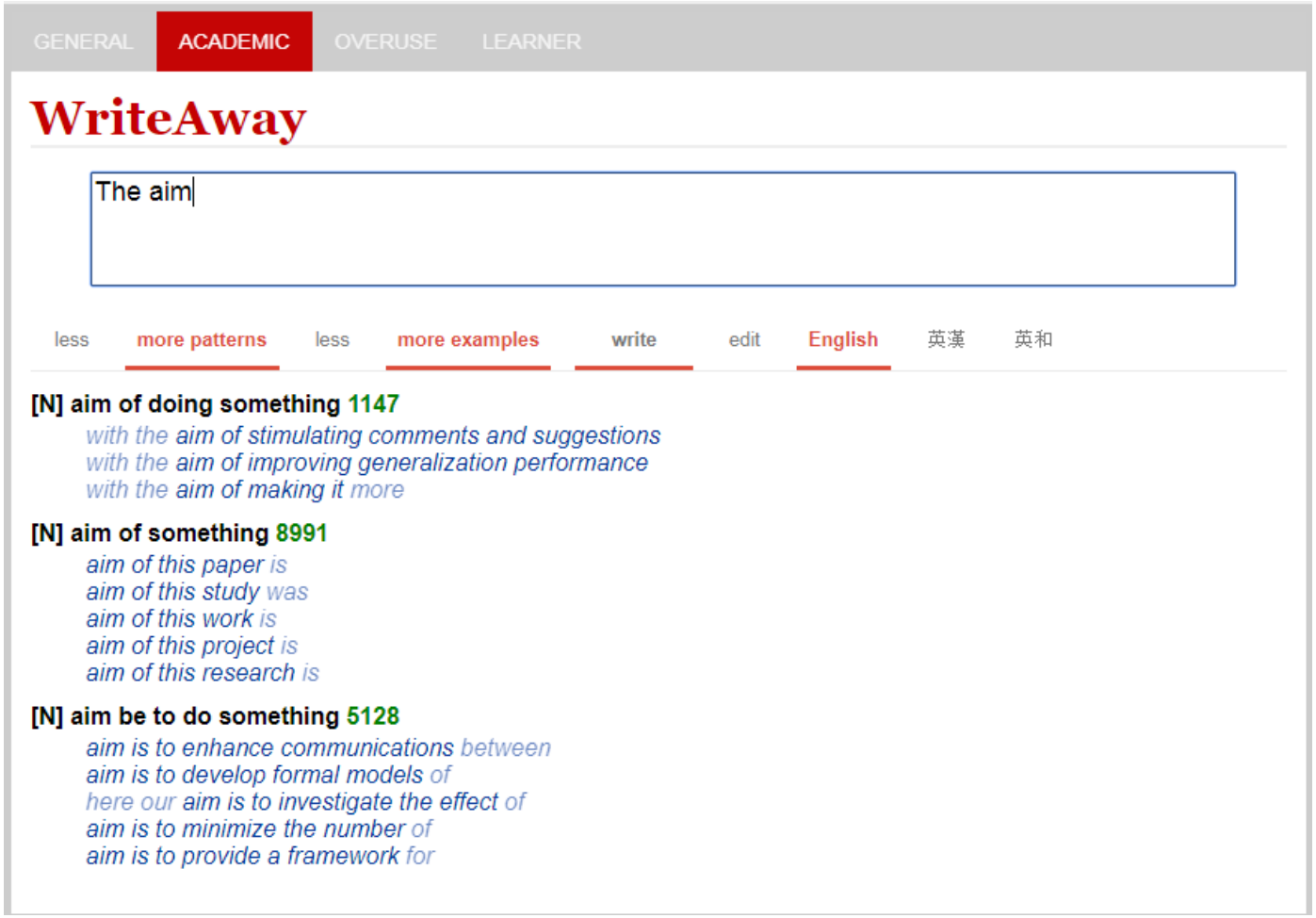
Existing real-time help with writing

Grammarly https://www.grammarly.com/



Existing real-time help with writing

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



Existing real-time help with writing

In summary

- Some truly innovative tools for real-time help during writing

But some limitations

- Mostly reactive feedback
- Feedback can be
    - prescriptive
    - distracting
    - intrusive
    - annoying

What's different about Collocaid?

- Will help EAP users with collocations in real time
- Raising awareness of collocation problems

A small increase of temperature

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

What's different about Collocaid?

- 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 +
- 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

What's different about Collocaid?

- 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

What's different about Collocaid?

Collocaid
Find the right words you need

- Customizable in more than one way

Research one as a result of…

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

Hide

Restore hidden prompts

- ☐ research
- ☐ hypothesis
- ☐ analyse

What's different about Collocaid?

- Prompts will be kept to a minimum, so that
  - they do not occupy too much screen space
  - users don't have to read lengthy explanations
- Metalanguage will be avoided, so that users don't have to know or decypher grammatical terminology

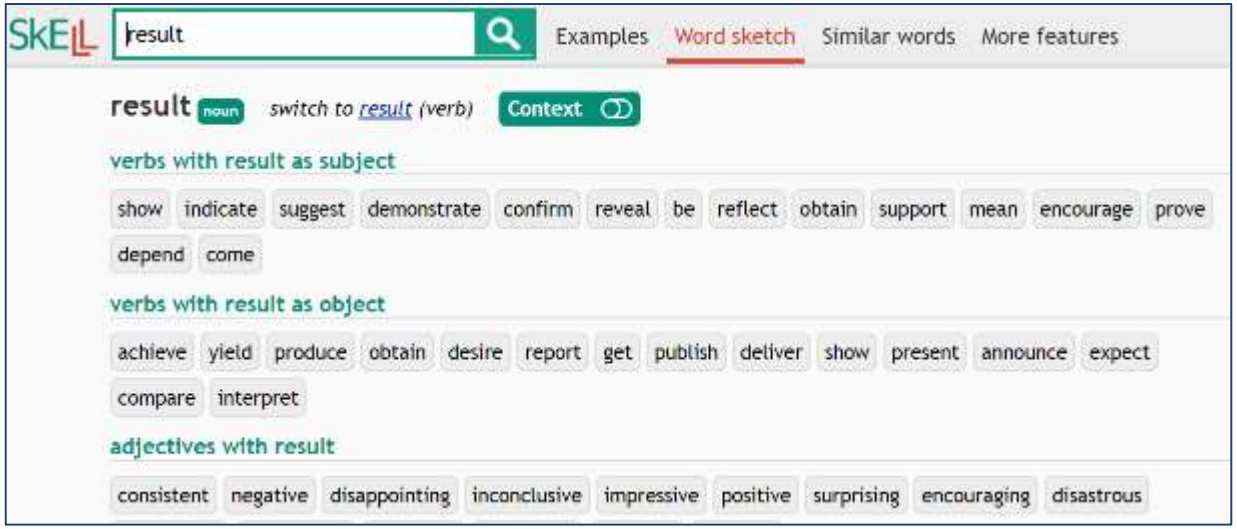
- Lexicographic data will be curated so that
  - we provide just the right amount
    - less is more
  - users don't get distracted from their writing with
    - irrelevant or misleading information
    - an overload of information

What's different about Collocaid?

Collocaid
Find the right words you need

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

  The results of the…



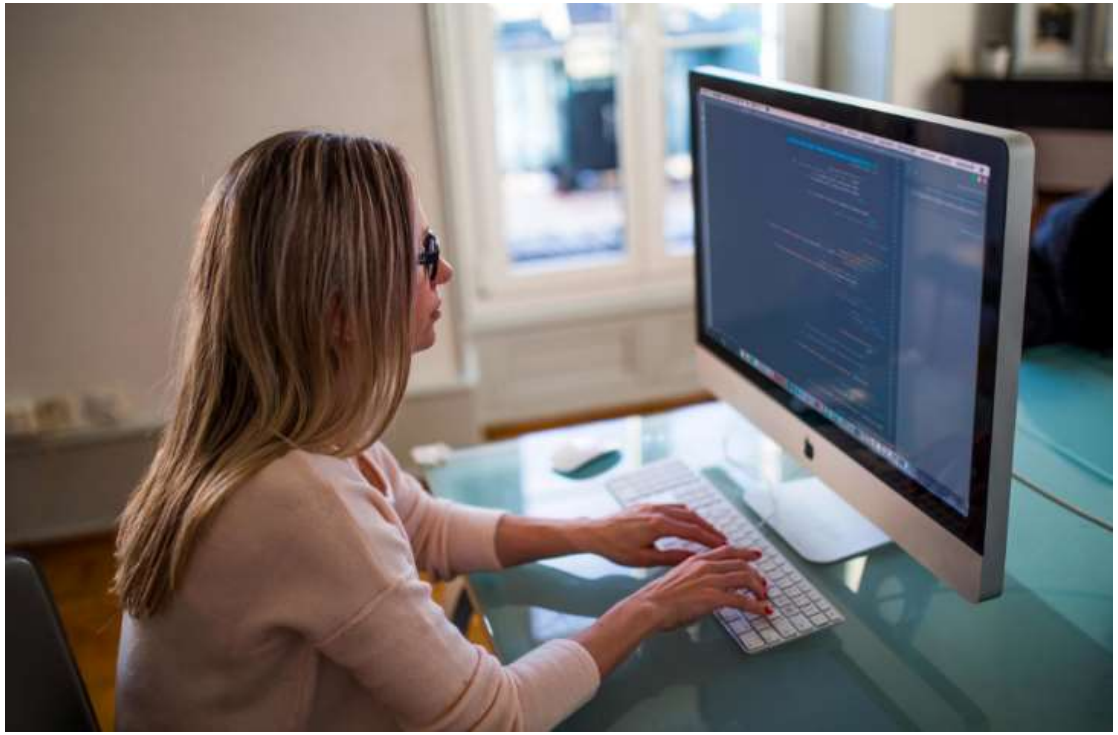
- User logs feed back to Collocaid database

What's different about Collocaid?

Collocaid

Find the right words you need

- Usability testing will be a priority



Compilation of 500 maximally useful collocation bases

- Academic Vocabulary List (AVL) (Gardner & Davies 2014)
  - Data from COCA_academic
  - Expert writing
- Cross-disciplinary AVL use in student writing (Durrant 2016)
- Academic Keyword List (Paquot 2010)
  - General English words frequently used in EAP
  - *increase, study…*

UNIVERSITY OF SURREY

COLLOCAID

Find the right words you need

Selection of strong collocates from EAP corpora

- Pearson International Corpus of Academic English, 37M
- Oxford Corpus of Academic English, 85M

Sketch Engine (Kilgariff et al. 2004)

- **datum** (noun) PICAE 2010

object_of	3,012	23.55
collect +	244	10.67
gather	83	9.37
analyse	74	8.96
interpret	65	8.59

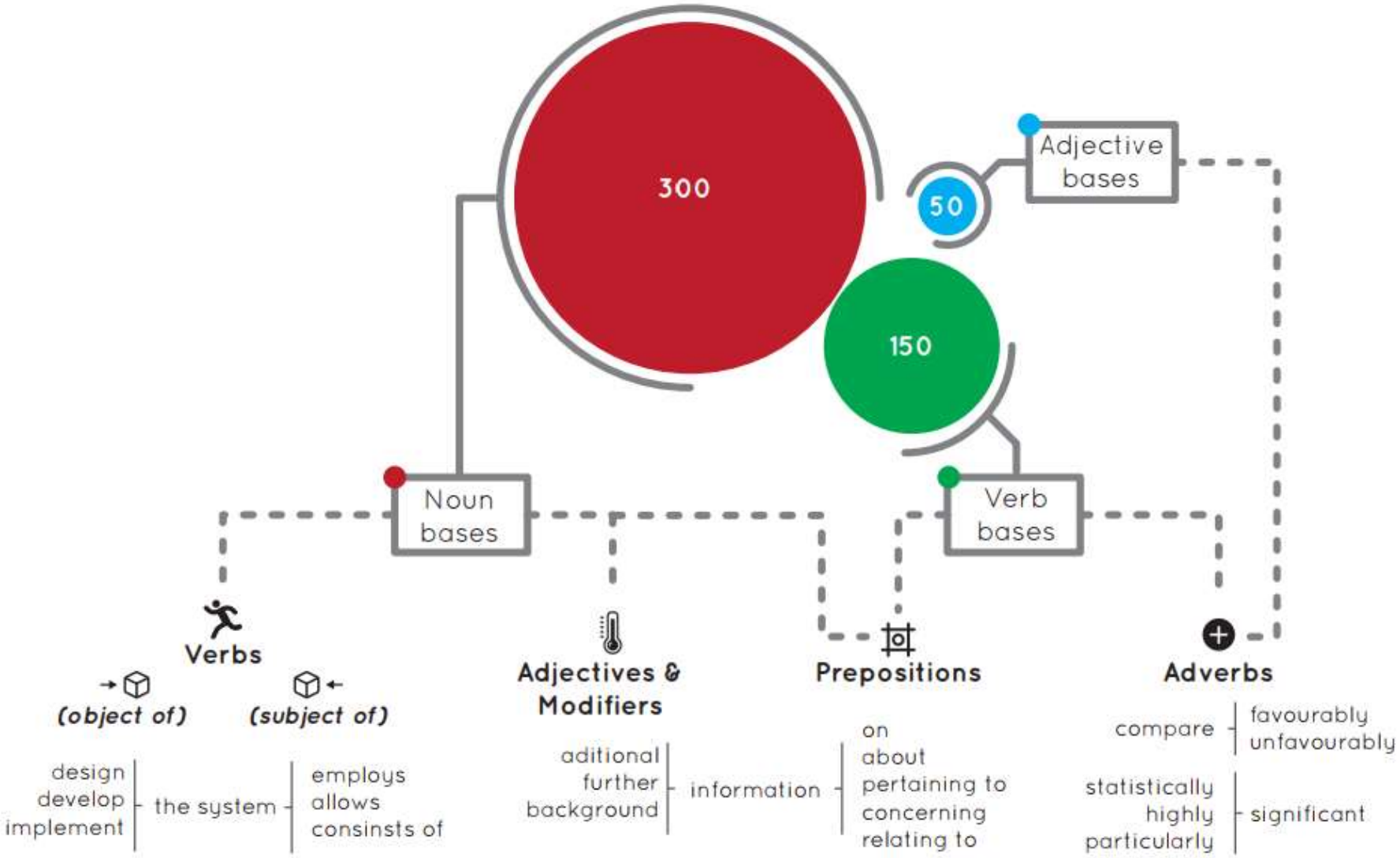
analyse (verb) PICAE 2010

AVP	202	10.25
critically	25	9.79
statistically	8	9.03
carefully	9	7.28
separately	3	6.99

significant (adjective) PICAE 2010

AVP	1,489	19.00
statistically +	120	11.10
most +	406	9.07
highly	69	8.45
particularly	67	8.39

Methodology



UNIVERSITY OF SURREY

COLLOCAID
Find the right words you need

Selection of good corpus examples from PICAE & OCAE

- Not too long, not too difficult
  - screen space
- Three examples of each (Frankenberg-Garcia 2014)
- Collocation highlighted in examples (Dziemianko 2014)

- **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

At the same time…

Research problematic EAP collocations

- Learner corpora
  - BAWE, ICLE, Open Cambridge Learner Corpus
- Existing resources (“get it right” & “error warning” boxes)
- Crowdsourced feedback (collocaid.uk)

At the same time…

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

Seek expert feedback from colleagues

Develop a prototype to test with EAP users

- build a prototype +
- a prototype featuring +
- an experimental prototype +
- a prototype of something +

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.

Carduner, J. (2003). Productive dictionaries skills training: What do language learners find useful? *Language Learning Journal*, *28*(1), 70–76.

Conklin, K., & **Schmitt, N.** (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, *32*, 45–61.

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.

Ellis, N., Simpson-Vlach, R., & **Maynard, C.** (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396.

Frankenberg-Garcia, A. (1999). Providing student writers with pre-text feedback. *ELT Journal*, *53*(2), 100–106.

Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, *18*(3), 335–355.

References

Frankenberg-Garcia, A. (2011). Beyond L1-L2 Equivalents: Where do Users of English as a Foreign Language Turn for Help? *International Journal of Lexicography*, *24*(1), 97–123.

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

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

Granger, S., & **Paquot, M.** (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica*, *31*(1), 118–141.

Hoey, M. (2005). *Lexical priming*. London: Routledge.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24–44.

Hyland, K. (2006)**.** *English for Academic Purposes: An Advanced Resource Book. London and New York: Routledge.*

Jenkins, J. (2014*). English as a Lingua Franca in the International University: The Politics of Academic English Language Policy*. Abingdon: Routledge.

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: 7-36**.**

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

Laufer, B., & **Waldman, T.** (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, *61*(2), 647–672.

References

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.

Lew, R., & **De Schryver, G.-M**. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography*, *27*(4), 341–359. https://doi.org/10.1093/ijl/ecu011

Nattinger, J., & **DeCarrico, J.** (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

Pawley, A., & **Syder, F. H.** (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike competence. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191– 227). Harlow: Longman.

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.

Stark, M. P. (1990). *Dictionary Workbooks: A Critical Evaluation of Dictionary Workbooks for the Foreign Language Learner*. Exeter: Exeter University Press.

Simpson-Vlach, R., & **Ellis, N.** (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, *31*, 487–512.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.