

Ana Frankenberg-Garcia
University of Surrey, UK
a.frankenberg-garcia@surrey.ac.uk

Robert Lew
Adam Mickiewicz University, Poland
rlew@amu.edu.pl

Jonathan C. Roberts
Bangor University, UK
j.c.roberts@bangor.ac.uk

Geraint P. Rees
University of Surrey, UK
g.rees@surrey.ac.uk

Javier Pereda
Bangor University, UK
j.pereda@bangor.ac.uk

Aims

- To research the most pervasive problems involving grammatical and lexical collocations encountered by writers of academic English and the collocations which they might find most useful
- To investigate how lexicographic data on collocation can be optimally integrated with digital writing environments

Identifying problematic collocations for users

Learner corpora

- BAWE (texts by university students in the UK)
- Cambridge Academic English Corpus (CAC) (includes writing from students at US and UK institutions)
- Cambridge Learner Corpus (CLC) (with texts from past First Certificate, Advanced Certificate and Proficiency papers)
- International Corpus of Learner English (Br-ICLE and P-ICLE)

Limitations

- BAWE and CAC may not disclose sufficient data on problematic collocations
 - The texts from CLC and ICLE might not be considered to be academic texts
- We nevertheless feel that inspecting how core academic collocation bases are used in these corpora can help us obtain a measure of recurrent problems shared by less proficient writers

Existing Literature

- 'Get it right' boxes in the Macmillan English Dictionary (Rundell, 2007)
- Error warning boxes in EAP textbooks e.g. Academic Vocabulary in Use (McCarthy & O'Dell, 2008) and scholarly work on miscollocations.

Crowdsourced Feedback



Scan the QR code or visit www.collocaid.uk to report collocation errors and join the debate

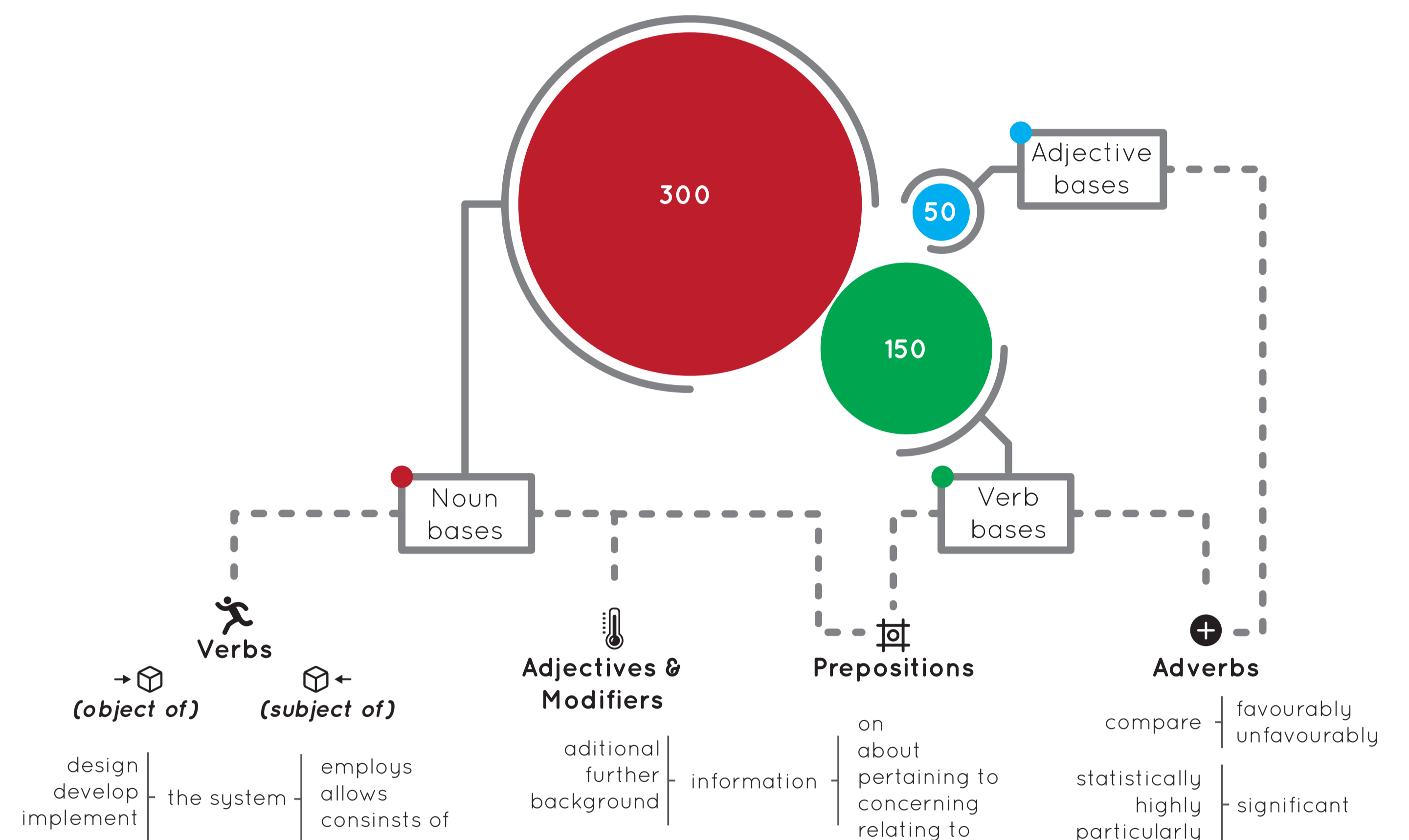
Identifying the base

- Academic Vocabulary List (AVL) (Gardner & Davies, 2014)
- Cross-disciplinary AVL use in student writing (Durrant, 2016)
- At least 500 maximally-useful bases (top AVL & problematic for learners)

Identifying salient collocations

Rationale

When researching what collocation prompts will be useful to writers, both lexical and grammatical collocations will be considered (e.g. *randomly selected* and *depending on*). The exact types of prompts we intend to offer writers will depend on the logical collocational paradigms they evoke:



Tools and resources

Sketch Engine (Kilgarriff et al., 2004)
Word Sketch (occurrence & logDice)

Pearson International Corpus of Academic English (Ackermann et al., 2010)

Oxford Corpus of Academic English (OCAE)

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

Integrating data with text editors

The objective of the study was to determine the effectiveness of advice from medical professionals to overweight men. One hundred participants were randomly assigned to treatment and control groups. Healthcare professionals gave advice to men in the treatment groups. At one-year follow-up, when analysed according to intention to treat, the treatment group reported significant weight loss compared with the control group ($p < 0.05$). Open-ended questionnaires were administered to ascertain the influencing factors. Successful weight loss depends of factors such as motivation, diet, and exercise.

MISCOLLOCATION PROMPTS → depend on

COLLOCATION SUGGESTIONS

- provide advice
- offer advice
- dispense advice

EXAMPLE SENTENCES the imperial guides offered advice on such problems they are incredibly supportive and offer constructive advice we are always here to offer advice

Your suggestions

What learner corpora could we use?

What other learner data might be helpful?

What target corpora could we consult?

What information on collocation should we provide users?

How can we integrate this information with text editors?