

Ústav Českého národního korpusu – Czech National Corpus Institute

Producing a dictionary of collocations

Michael Rundell
Macmillan Dictionaries and Lexicography MasterClass

Outline

- Why a collocations dictionary?
 - for receptive use
 - for productive use
- Questions for a collocations dictionary
 - which definition of collocation?
 - what should it include?
- A look at MCD

What is collocation?

- It is not so much the words of English nor the grammar of English that makes English difficult...The vague and undefined obstacle to progress ... consists for the most part in the existence of so many odd comings-together-of words
- Harold Palmer 1933

But why a collocations dictionary?

- What's wrong with
 - regular learners' dictionaries?
 - online sources, such as
 - Leo <http://dict.leo.org>
 - Frequency lists www.wordfrequency.info/
 - Just the Word
<http://193.133.140.102/JustTheWord/>
- Most of the information is there, most of it is free, but ...

What makes collocation so important?

- ❑ Link between *frequency* (in a corpus), *typicality*, and *naturalness*
- ❑ In productive mode: a key to fluency and naturalness
- ❑ In receptive mode, essential to meaning: you really do ‘know a word by the company it keeps’

Collocation and meaning

- **Sick** has several meanings: which one does the speaker intend?
 - *I had never eaten oysters before, and I became violently sick.*
 - *Volunteers are trained to provide care for chronically sick people in their homes.*
 - *We are all heartily sick of their negative attitude.*
- The collocates (adverb) is the clue

Economic/economical: what's the difference?

- Look at the nouns which these adjectives most frequently modify
- These are its collocates
- What do they tell us about the meaning of the two adjectives?
- See also data in MED
 - 'Get it Right' box at entry for *economical*
 - 'Collocation box' at entry for *economic*

economic *(adjective)* LEXMCI freq =

AVP	3248	0.3	N_premod	184427	3.6
more	416	3.19	growth	12917	9.75
purely	266	8.35	development	10030	7.94
most	264	3.27	policy	4578	7.17
well	193	2.41	activity	4544	7.13
further	151	4.63	benefit	3169	7.34
closer	110	7.0	crisis	3163	8.33
only	109	1.5	reform	3133	8.27
about	108	2.74	impact	2151	7.22
just	75	0.93	condition	2077	6.37
faster	52	5.85	sanction	1903	8.06
primarily	49	4.81	system	1847	4.73
then	42	0.25	problem	1732	5.36
particularly	39	2.72	issue	1732	5.31
especially	34	3.37	performance	1603	6.23
less	31	2.37	recovery	1516	7.55
even	28	0.54	interest	1482	5.8
mainly	27	3.28	situation	1470	6.43

economical *(adjective)* LEXMCI fre

Constructions	N_premod	3505	2.0
Vinf_to	432	3.1	
it_Vinf_to	131	55.8	

AVP	2271	6.3
more	904	4.31
most	591	4.43
very	230	2.25
extremely	43	3.49
highly	37	2.81
less	31	2.37
quite	19	0.95
relatively	14	2.28
fairly	12	2.21
longer	11	1.29
particularly	11	0.9
yet	11	0.35
reasonably	9	2.88
somewhat	9	2.16

way	350	2.65
solution	123	3.28
use	100	1.92
method	80	2.37
alternative	67	4.22
price	67	1.91
option	59	2.1
car	56	1.42
mean	50	3.6
manner	35	3.09
repair	34	3.52
engine	33	1.96
route	28	1.63
choice	28	1.55
cost	24	0.17
rate	24	0.1
benefit	22	0.65
administration	20	1.98

Two views of collocation

The statistical view

a measure of the frequency with which two or more words co-occur (cf. Cermak 2006: 'purely surface, computational' approach)

■ *of the*: the most frequent 'bigram' in English

■ the most frequent adverb+adjective combinations include:

very good, more likely, quite happy, most important

Is this interesting/useful information?

Two views of collocation

□ An onomasiological/functional view

- how do I say ‘do/make’ with this noun

- e.g. **she made a serious crime* ?

- how do I say ‘very/extremely’ with this adjective

- e.g. *it was bitterly cold?*

- how do I say ‘strongly’ with this verb

- e.g. *I distinctly (not **strongly*) remember*

What should a 'collocations dictionary' include?

- Language data isn't a problem
 - very large corpora available (2 billion words)
 - intelligent software (e.g. Word Sketches), with saliency measures: finds all frequent collocates

impact (noun) LEXMCI freq = 217441

Constructions		
PP_Ving	5186	3.3
that_0	5156	1.0

PP_Ving	5186	3.3
upon	74	3.24
on	1470	2.06
whilst	14	1.84
while	28	1.06
of	2648	0.88
without	33	0.76
since	18	0.57
by	297	0.13

PP_NP_Ving	1630	2.9
upon	31	2.01
on	632	0.84

PP_cl_wh	920	1.8
upon	24	1.64
on	579	0.71

object_of	91820	3.3
minimise	2022	9.03
assess	3417	8.92
evaluate	1180	7.84
reduce	3091	7.7
mitigate	492	7.35
examine	1167	7.27
maximise	503	7.14
measure	846	7.07
lessen	328	6.76
investigate	717	6.71

N_premod	17772	0.2
assessment	3813	8.48
crater	182	7.92
resistant	46	6.11
ionisation	37	5.96
resistance	165	5.69
velocity	61	5.68
wrench	27	5.47
excitation	29	5.47
oscillator	29	5.44
implication	150	5.41

N_mod	16199	0.3
asteroid	141	7.71
equality	491	7.68
meteorite	100	7.41
climate	470	6.68
meteor	50	6.36
comet	47	5.93
knock-on	34	5.93
noise	201	5.71

SUPPREP	39030	0.9
about	3305	5.75
despite	123	4.55
upon	145	4.04
although	122	3.89
whether	113	3.84
on	4799	3.76
without	216	3.4
into	727	3.36

AJ_premod	88765	3.4
environmental	8097	9.94
negative	3707	9.52
adverse	2134	9.18
positive	4084	8.97
significant	5284	8.92
potential	3340	8.73
visual	1470	8.08
profound	801	7.85

PP_PP	44155	5.0
of_on	16079	10.18
on_of	8757	8.21
of_upon	490	8.19
on_in	1838	7.88
upon_of	486	7.68
on_as	520	7.51
on_than	275	7.44
on_through	162	6.71

AVP_post_mod
negatively
whatsoever
upon
adversely
positively
disproportionate
directly
locally

But which words do you include?

- Not obvious – wide range of answers in existing collocations dictionaries
- No established model (unlike general ‘advanced learner’s dictionaries’)
 - BBI: non-corpus based, follows Mel’cuk’s categories (broadly)
 - LTP: non-corpus-based, no clear methodology
 - OCD: see next slide
- Depends on how you define collocation

A descriptive or productive approach?

- Oxford Collocations Dictionary: *descriptive*
 - provides a complete record of how words combine
 - wide coverage: headwords include
 - *goalkeeper, seagull, admiral, masochism, flea, double bass, nervous breakdown, diarrhoea*
- Very detailed descriptions: e.g. *food*

'Collocationality' (Kilgarriff 2006): a more scientific approach

- Some words are more 'collocational' than others
- A spectrum, going from
 - *unmitigated, thorny, ulterior, amends, wreak*: a small number of very salient collocates
- to
 - *food, house, big, feel*: a large number of combinations, none very strong
- Interesting area is in the middle!

Macmillan Collocations Dictionary (MCD): a *productive* resource

- Selective:
 - words with medium collocationality: not too high (*wreak*), not too low (*house*)
 - special focus on academic and professional contexts
- 4500 headwords (less is more)
- Includes most items in the Academic Word List, e.g.
 - *analogy, argue, assume, assertion, benefit, coherent, compelling, conclude, contrast ...*

Structure: what's in an entry

- Main division: grammatical relations
- e.g. *analysis* (N)
 - v+N (*conduct, carry out, present ...*)
 - adj+N (*thorough, in-depth, preliminary ...*)
 - N+v (*confirm, demonstrate, suggest...*)
 - n+N (*cost-benefit, risk, content ...*)

Structure: what's in an entry

- Secondary division: semantic sets
- e.g. *advice*, v+N relation: separate sets for
 - ask for advice
 - give advice
 - follow advice
 - disregard advice [etc]
- Each set has definition, collocates, example(s)

Special features: notes on ...

□ Colligation

- *problem*: 'In these combinations, problem is always plural' (e.g. *marital*)
- *accept*: 'In these combinations, accept is usually passive' (e.g. *generally, widely*)

□ Alternatives: chunks not collocations

- *advice*, section on 'disregarding advice': 'If you disregard someone's advice, you can say you do something *against someone's advice*

Special features: sets with prepositions, infinitives

- V+with: *regard with* (respect, affection, contempt ...)
- N+for: *recipe for* (disaster, confusion, success ...)
- n+of+N: *regret* (a feeling, tinge..)
- ADJ+infinitive: *reasonable to* (assume, conclude, believe...)

Special feature: focus on the ‘ collocator’ not the ‘base’ (cf. Hausmann 1997)

□ Verbs like

- *exercise* (caution, restraint, discretion ...)
- *voice* (concern, anger, displeasure ...)
- *allocate* (resources, funding, budget ...)

□ Adjectives like

- *adverse* (effects, reaction, consequences ...)
- *arbitrary* (arrest, detention, imprisonment)
- *exhaustive* (analysis, review, survey ...)

MCD entry for implication

Collocates are
divided into
semantic groups,
with short definition
for each group

implication N

1 [usually plural] a possible effect or result

- adj+N important **considerable, enormous, important, profound, significant** *These differences have significant implications **for** growth and development.*
- ▶ negative **devastating, negative, serious** *This ruling could have devastating implications **for** thousands of people.*
- ▶ possible **possible, potential** *The potential legal implications of this approach are enormous.*
- ▶ affecting a large number of people or things **broad, far-reaching, wide, wide-ranging** *The legislation has far-reaching implications **for** property owners in the UK.*
- ▶ over a long period of time **long-term** *We need to carefully consider the long-term implications of development.*
- ▶ relating to a real situation **practical** *This report has practical implications **for** our work in the classroom.*
- ▶ relating to money **budgetary, economic, financial** *You should consider carefully the financial implications of studying for a higher degree.*

budget N

the money that a person, organization or government can spend

- v+N decide on and manage a budget **allocate, approve, set, agree, propose, prepare, manage** *Just like any other significant purchase, you set a realistic budget.*
- ▶ manage a budget successfully **balance, control, stay within, work within, keep within, keep to** *Last year we delivered our targets and balanced our budget. • Plan carefully to keep within budgets and deadlines.*
- ▶ spend a lot of or more than a budget **overspend, spend, exceed, blow, go over, run over, stretch** *It may not be necessary to blow the budget on special equipment. • The production had also gone well over budget. • Having to pay for new flights really stretched our budget.*
- ▶ change a budget **revise** *We found it necessary to revise the budget halfway through the financial year.*
- ▶ increase a budget **increase** *The project will fail unless the budget is increased.*
- ▶ cut a budget **slash, cut, reduce** *Last year we cut the budget by 50 percent.*