OXFORD
UNIVERSITY PRESS

From academic corpus to EAP dictionary

Diana Lea

BALEAP PIM, Coventry University
21 June 2014

Help with academic writing

- Using appropriate language
- Collocations and synonyms
- Planning and structuring assignments
- Presenting an argument
- Using sources correctly

From corpus to dictionary

- Composition of the corpus
- Creation of dictionary entries
- Insights into academic vocabulary
- Evaluation of the dictionary

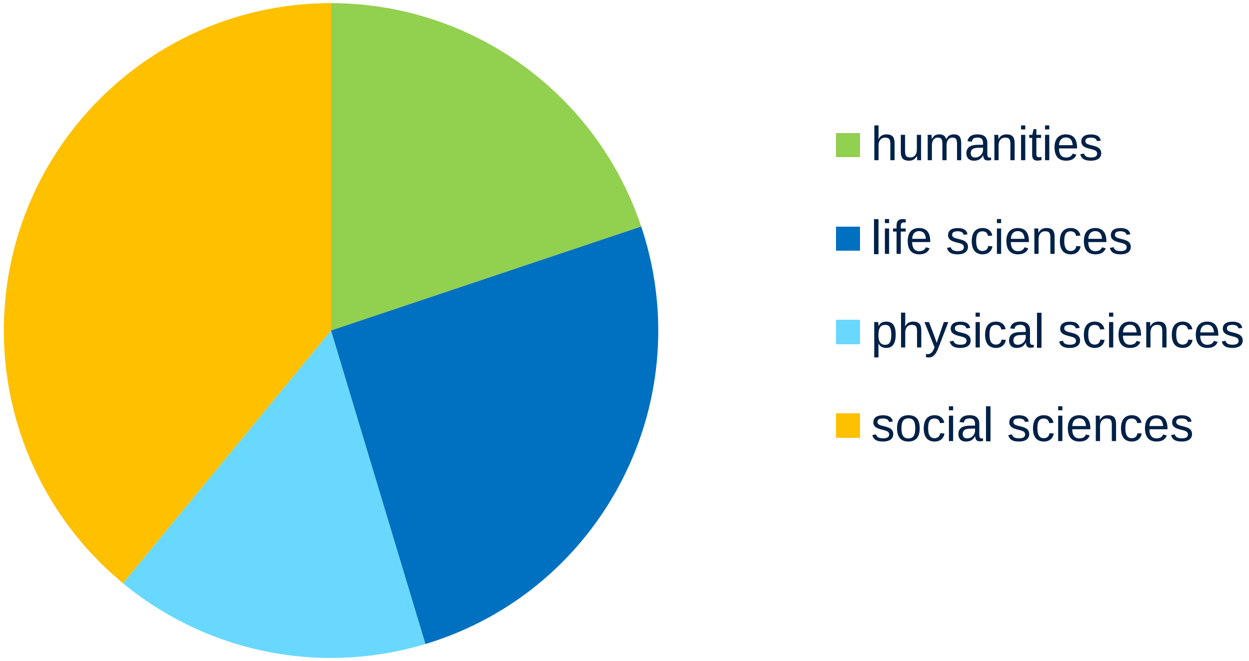
Who needs an EAP dictionary?

- Students on English-medium degree courses
- Students on foundation or pre-sessional courses
- Students from B1+ level upwards
- Students of all subjects (business, medicine, engineering, computer science …)

The academic corpus

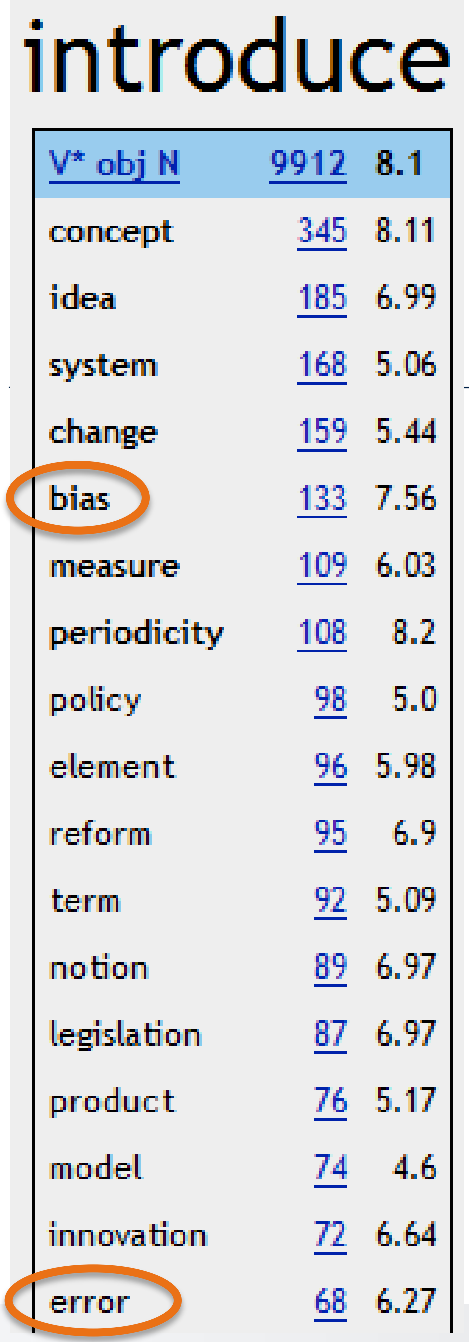
- Oxford Corpus of Academic English
- 85 million words
- Higher Education textbooks and academic journals
- 4 subcorpora based on subject areas
- All example sentences are corpus-based

The academic corpus



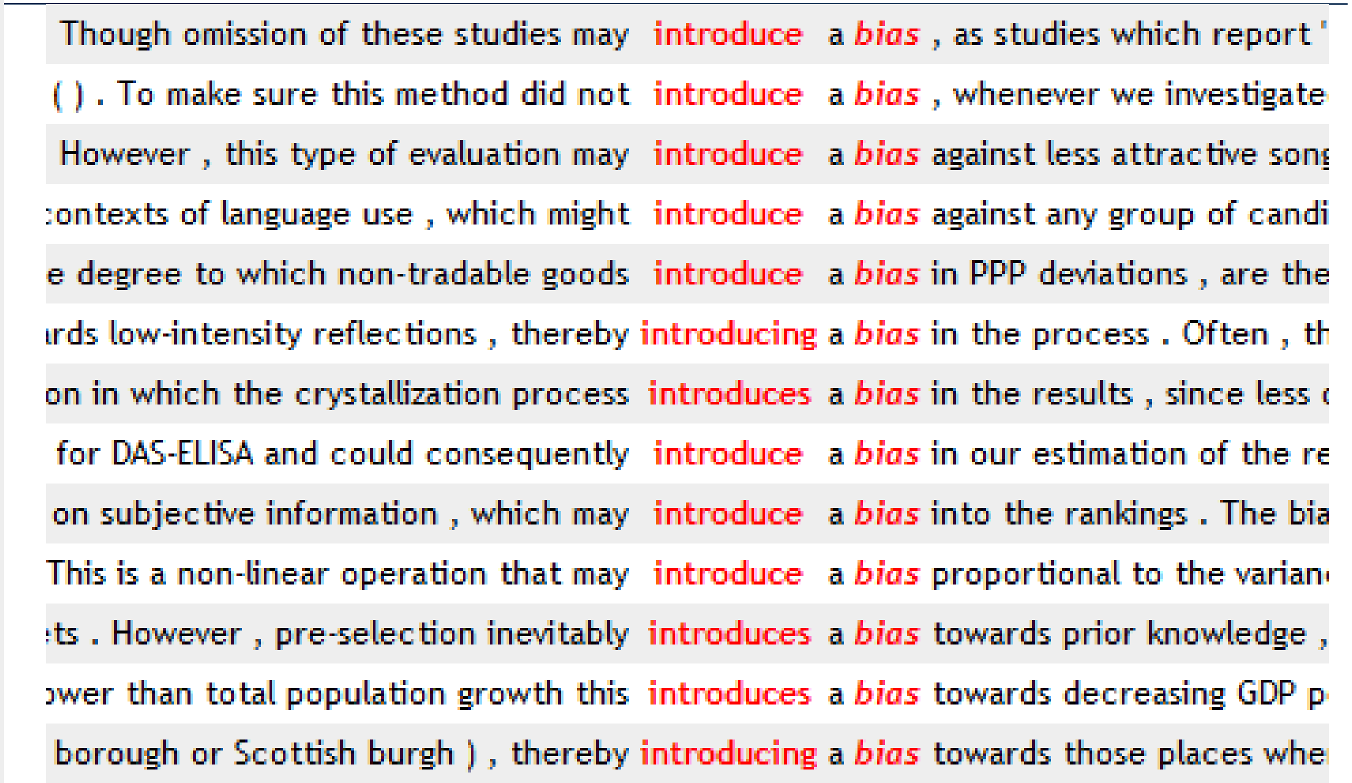
What is the main vocabulary focus?

- General academic vocabulary
- Academic Word List: 570 word families – over 1,600 words
- Defining vocabulary – 2,300 words
- Synonyms, opposites and collocations of AWL words
- 4 subject area word lists of 200-300 words each



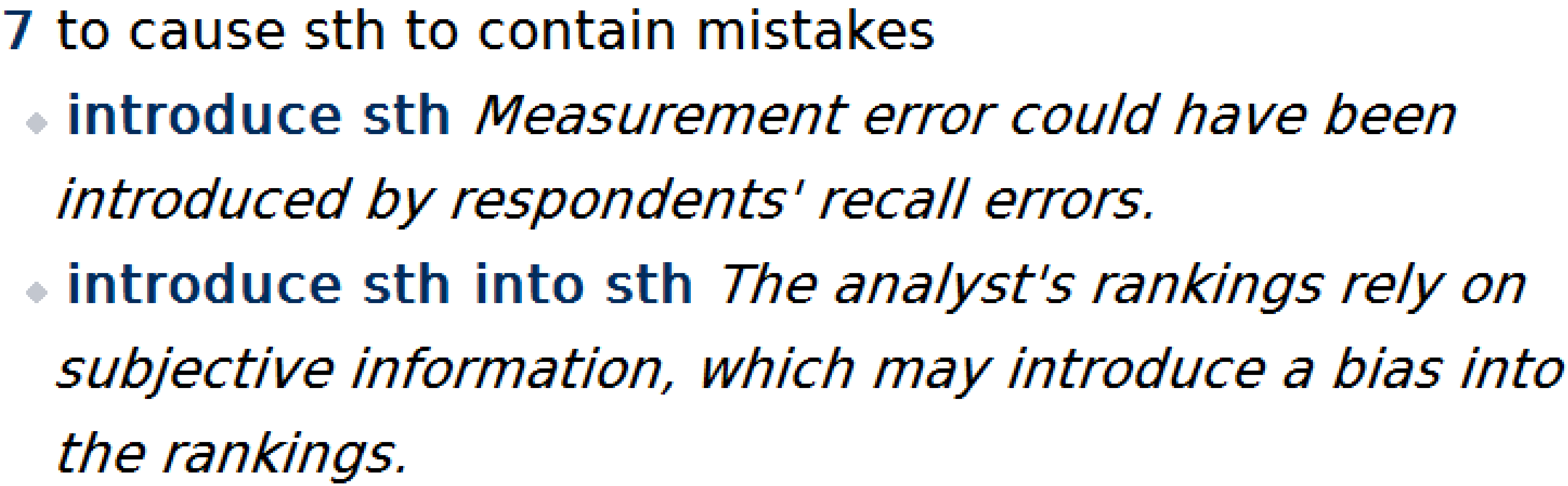
introduce

V* obj N	9912	8.1
concept	345	8.11
idea	185	6.99
system	168	5.06
change	159	5.44
bias	133	7.56
measure	109	6.03
periodicity	108	8.2
policy	98	5.0
element	96	5.98
reform	95	6.9
term	92	5.09
notion	89	6.97
legislation	87	6.97
product	76	5.17
model	74	4.6
innovation	72	6.64
error	68	6.27



Though omission of these studies may introduce a *bias* , as studies which report '
( ) . To make sure this method did not introduce a *bias* , whenever we investigate
However , this type of evaluation may introduce a *bias* against less attractive song
ontexts of language use , which might introduce a *bias* against any group of candi
e degree to which non-tradable goods introduce a *bias* in PPP deviations , are the
rds low-intensity reflections , thereby introducing a *bias* in the process . Often , th
on in which the crystallization process introduces a *bias* in the results , since less
for DAS-ELISA and could consequently introduce a *bias* in our estimation of the re
on subjective information , which may introduce a *bias* into the rankings . The bia
This is a non-linear operation that may introduce a *bias* proportional to the varian
ts . However , pre-selection inevitably introduces a *bias* towards prior knowledge ,
ower than total population growth this introduces a *bias* towards decreasing GDP p
borough or Scottish burgh) , thereby introducing a *bias* towards those places whe

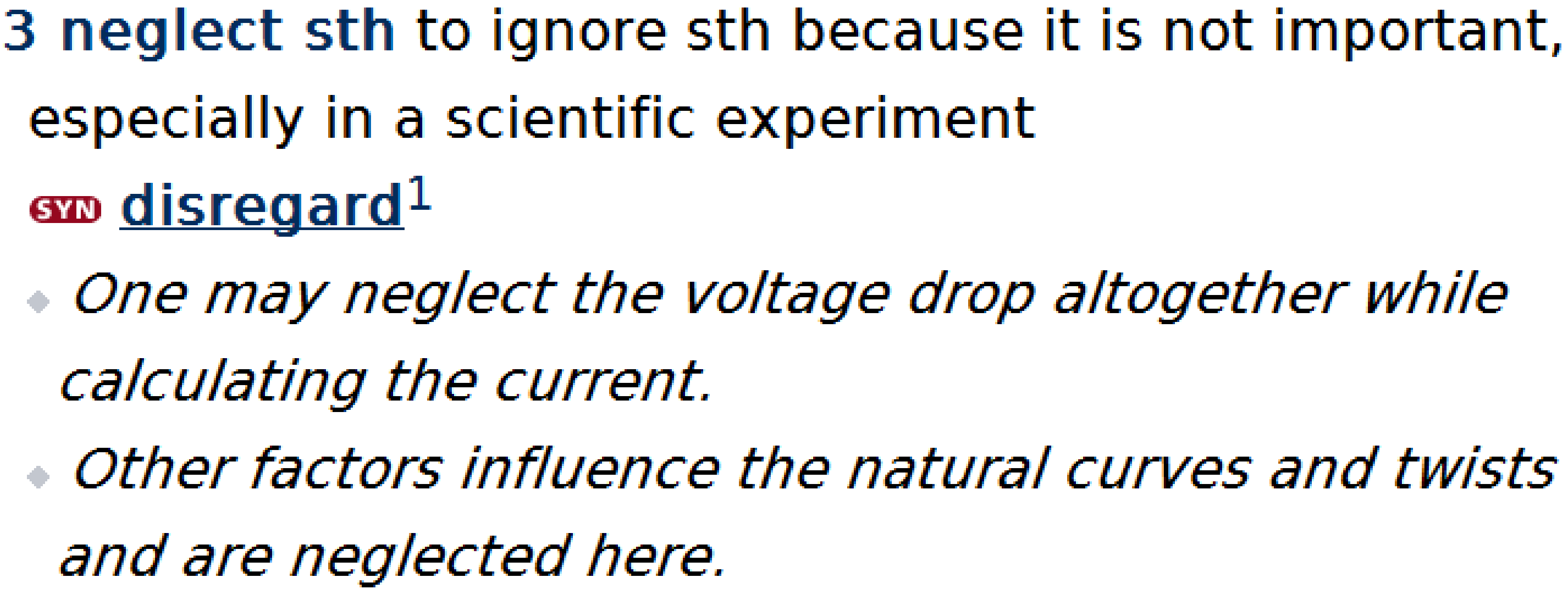
Creating the entries



7 to cause sth to contain mistakes

- **introduce sth** *Measurement error could have been introduced by respondents' recall errors.*
- **introduce sth into sth** *The analyst's rankings rely on subjective information, which may introduce a bias into the rankings.*

Creating the entries



3 **neglect sth** to ignore sth because it is not important, especially in a scientific experiment

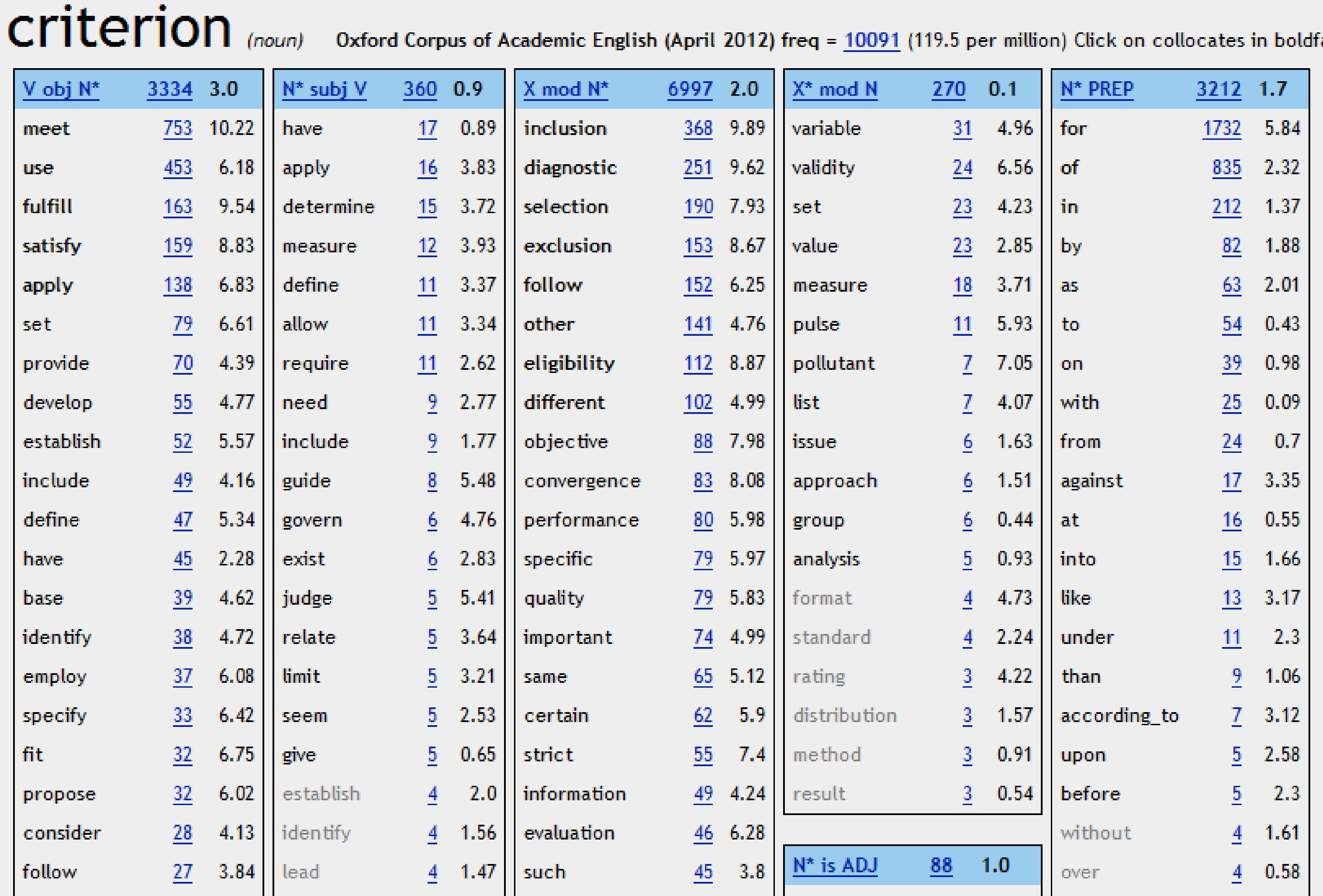
SYN disregard^{1}

◆ *One may neglect the voltage drop altogether while calculating the current.*

◆ *Other factors influence the natural curves and twists and are neglected here.*

Creating the entries

- meaning
- grammar
- complementation patterns
- collocations
- synonyms
- functions



criterion *(noun)*

Oxford Corpus of Academic English (April 2012) freq = 10091 (119.5 per million) Click on collocates in boldfa

V obj N*	3334	3.0
meet	753	10.22
use	453	6.18
fulfill	163	9.54
satisfy	159	8.83
apply	138	6.83
set	79	6.61
provide	70	4.39
develop	55	4.77
establish	52	5.57
include	49	4.16
define	47	5.34
have	45	2.28
base	39	4.62
identify	38	4.72
employ	37	6.08
specify	33	6.42
fit	32	6.75
propose	32	6.02
consider	28	4.13
follow	27	3.84

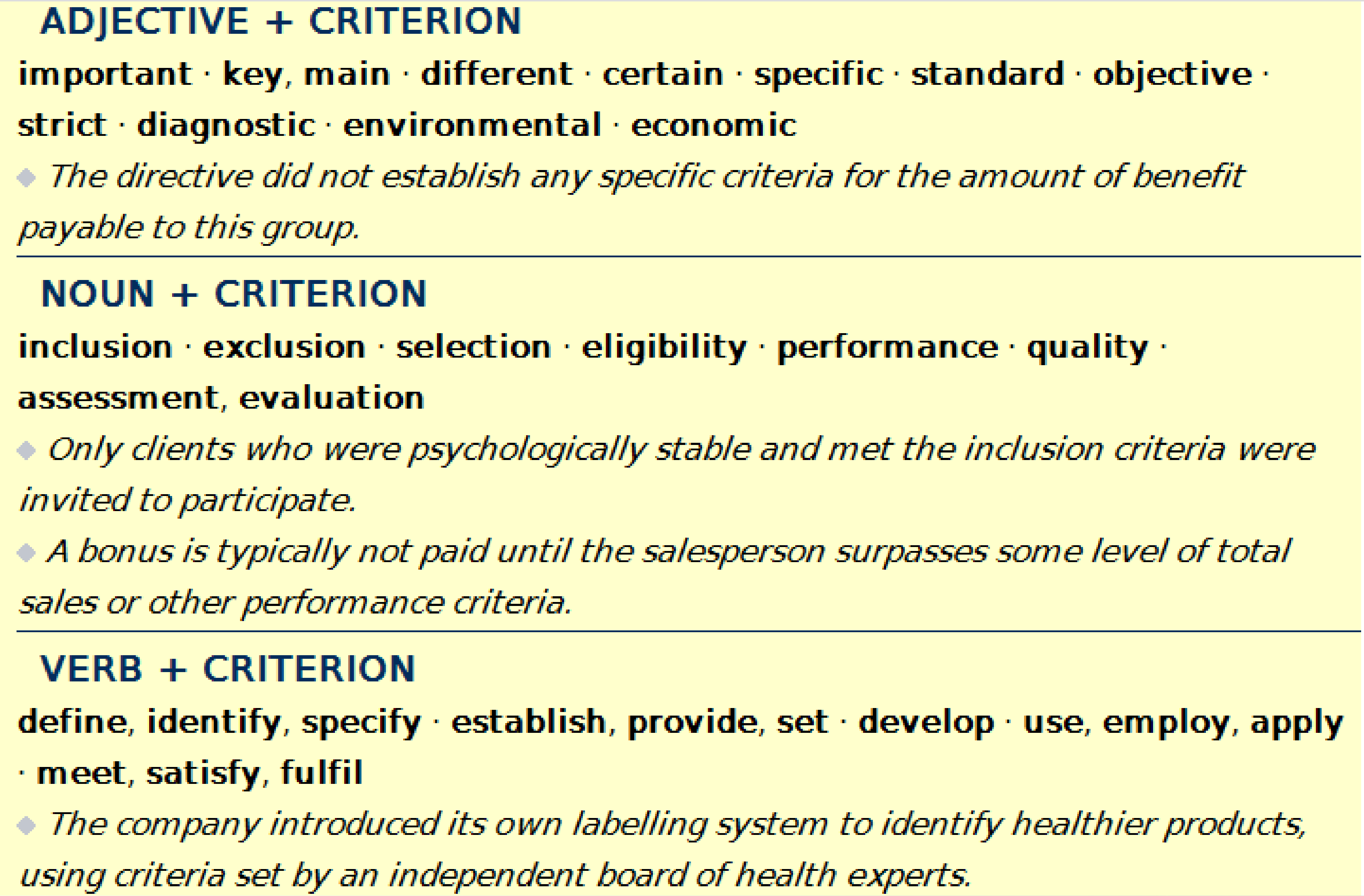
N* subj V	360	0.9
have	17	0.89
apply	16	3.83
determine	15	3.72
measure	12	3.93
define	11	3.37
allow	11	3.34
require	11	2.62
need	9	2.77
include	9	1.77
guide	8	5.48
govern	6	4.76
exist	6	2.83
judge	5	5.41
relate	5	3.64
limit	5	3.21
seem	5	2.53
give	5	0.65
establish	4	2.0
identify	4	1.56
lead	4	1.47

X mod N*	6997	2.0
inclusion	368	9.89
diagnostic	251	9.62
selection	190	7.93
exclusion	153	8.67
follow	152	6.25
other	141	4.76
eligibility	112	8.87
different	102	4.99
objective	88	7.98
convergence	83	8.08
performance	80	5.98
specific	79	5.97
quality	79	5.83
important	74	4.99
same	65	5.12
certain	62	5.9
strict	55	7.4
information	49	4.24
evaluation	46	6.28
such	45	3.8

X* mod N	270	0.1
variable	31	4.96
validity	24	6.56
set	23	4.23
value	23	2.85
measure	18	3.71
pulse	11	5.93
pollutant	7	7.05
list	7	4.07
issue	6	1.63
approach	6	1.51
group	6	0.44
analysis	5	0.93
format	4	4.73
standard	4	2.24
rating	3	4.22
distribution	3	1.57
method	3	0.91
result	3	0.54

N* is ADJ	88	1.0

N* PREP	3212	1.7
for	1732	5.84
of	835	2.32
in	212	1.37
by	82	1.88
as	63	2.01
to	54	0.43
on	39	0.98
with	25	0.09
from	24	0.7
against	17	3.35
at	16	0.55
into	15	1.66
like	13	3.17
under	11	2.3
than	9	1.06
according_to	7	3.12
upon	5	2.58
before	5	2.3
without	4	1.61
over	4	0.58



ADJECTIVE + CRITERION

important · key, main · different · certain · specific · standard · objective · strict · diagnostic · environmental · economic

◆ *The directive did not establish any specific criteria for the amount of benefit payable to this group.*

NOUN + CRITERION

inclusion · exclusion · selection · eligibility · performance · quality · assessment, evaluation

◆ *Only clients who were psychologically stable and met the inclusion criteria were invited to participate.*

◆ *A bonus is typically not paid until the salesperson surpasses some level of total sales or other performance criteria.*

VERB + CRITERION

define, identify, specify · establish, provide, set · develop · use, employ, apply · meet, satisfy, fulfil

◆ *The company introduced its own labelling system to identify healthier products, using criteria set by an independent board of health experts.*

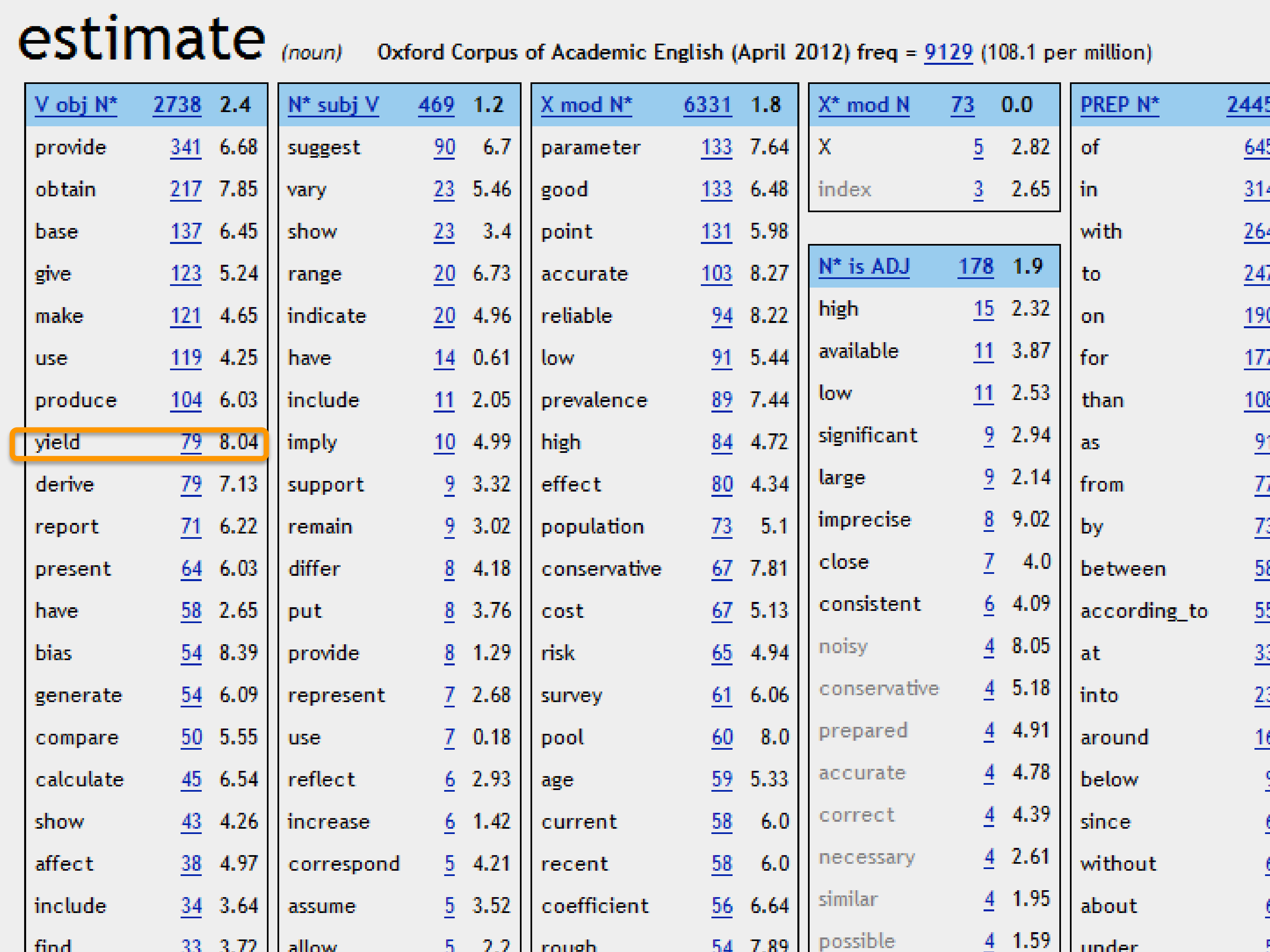
What makes a good dictionary example?

Examples need to be …

- comprehensible
- convincing
- typical
- transferable

What makes a good dictionary example?

- *Taylor makes the following argument: …*
- *This approach yields dramatically lower estimates.*
- *Several other factors played a role in the decision-making.*
- *The most persuasive argument against this idea comes from Foster (2009).*



estimate *(noun)* Oxford Corpus of Academic English (April 2012) freq = 9129 (108.1 per million)

V obj N*	2738	2.4
provide	341	6.68
obtain	217	7.85
base	137	6.45
give	123	5.24
make	121	4.65
use	119	4.25
produce	104	6.03
yield	79	8.04
derive	79	7.13
report	71	6.22
present	64	6.03
have	58	2.65
bias	54	8.39
generate	54	6.09
compare	50	5.55
calculate	45	6.54
show	43	4.26
affect	38	4.97
include	34	3.64
find	33	3.72

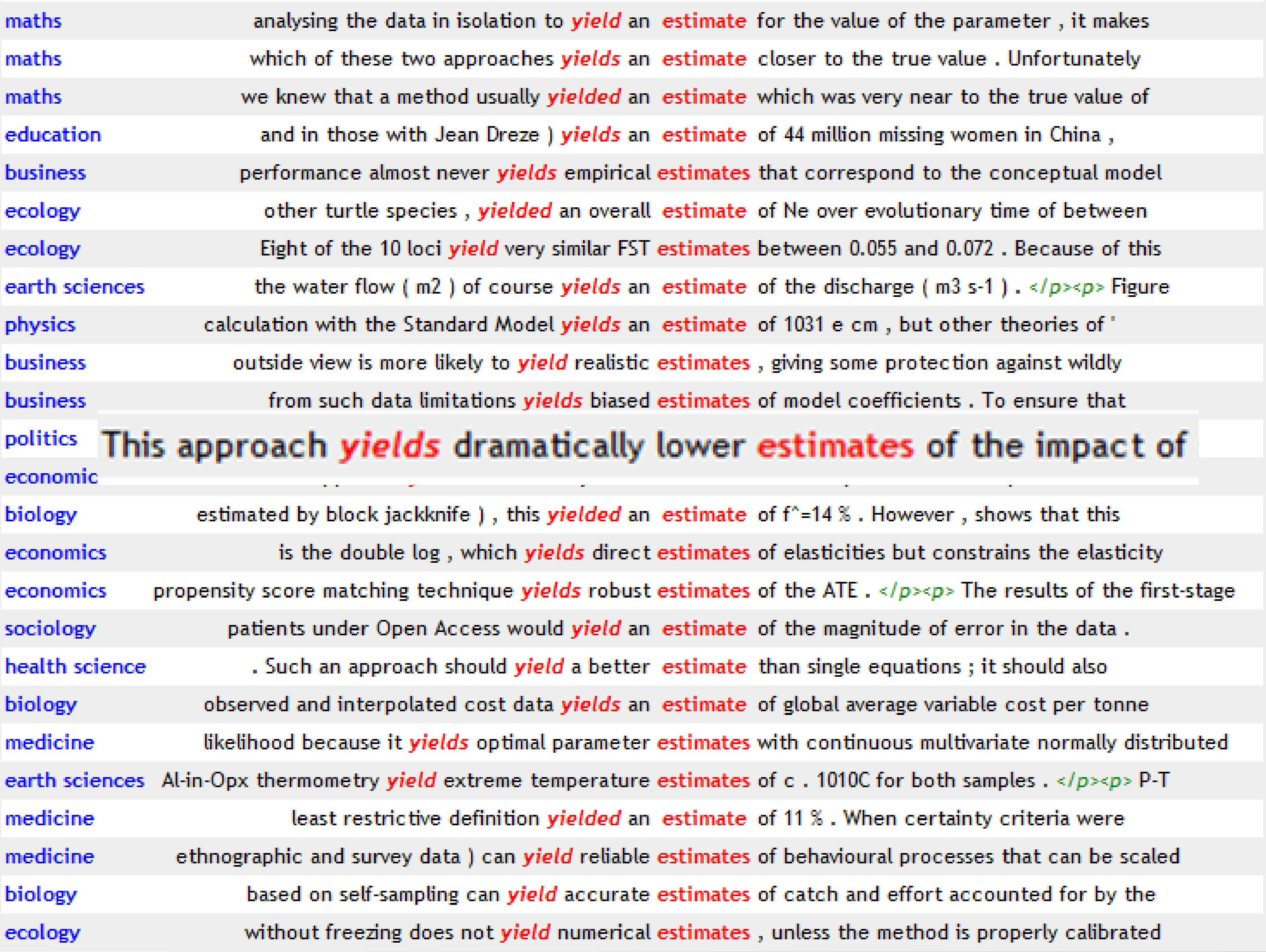
N* subj V	469	1.2
suggest	90	6.7
vary	23	5.46
show	23	3.4
range	20	6.73
indicate	20	4.96
have	14	0.61
include	11	2.05
imply	10	4.99
support	9	3.32
remain	9	3.02
differ	8	4.18
put	8	3.76
provide	8	1.29
represent	7	2.68
use	7	0.18
reflect	6	2.93
increase	6	1.42
correspond	5	4.21
assume	5	3.52
allow	5	2.2

X mod N*	6331	1.8
parameter	133	7.64
good	133	6.48
point	131	5.98
accurate	103	8.27
reliable	94	8.22
low	91	5.44
prevalence	89	7.44
high	84	4.72
effect	80	4.34
population	73	5.1
conservative	67	7.81
cost	67	5.13
risk	65	4.94
survey	61	6.06
pool	60	8.0
age	59	5.33
current	58	6.0
recent	58	6.0
coefficient	56	6.64
rough	54	7.89

X* mod N	73	0.0
X	5	2.82
index	3	2.65

N* is ADJ	178	1.9
high	15	2.32
available	11	3.87
low	11	2.53
significant	9	2.94
large	9	2.14
imprecise	8	9.02
close	7	4.0
consistent	6	4.09
noisy	4	8.05
conservative	4	5.18
prepared	4	4.91
accurate	4	4.78
correct	4	4.39
necessary	4	2.61
similar	4	1.95
possible	4	1.59

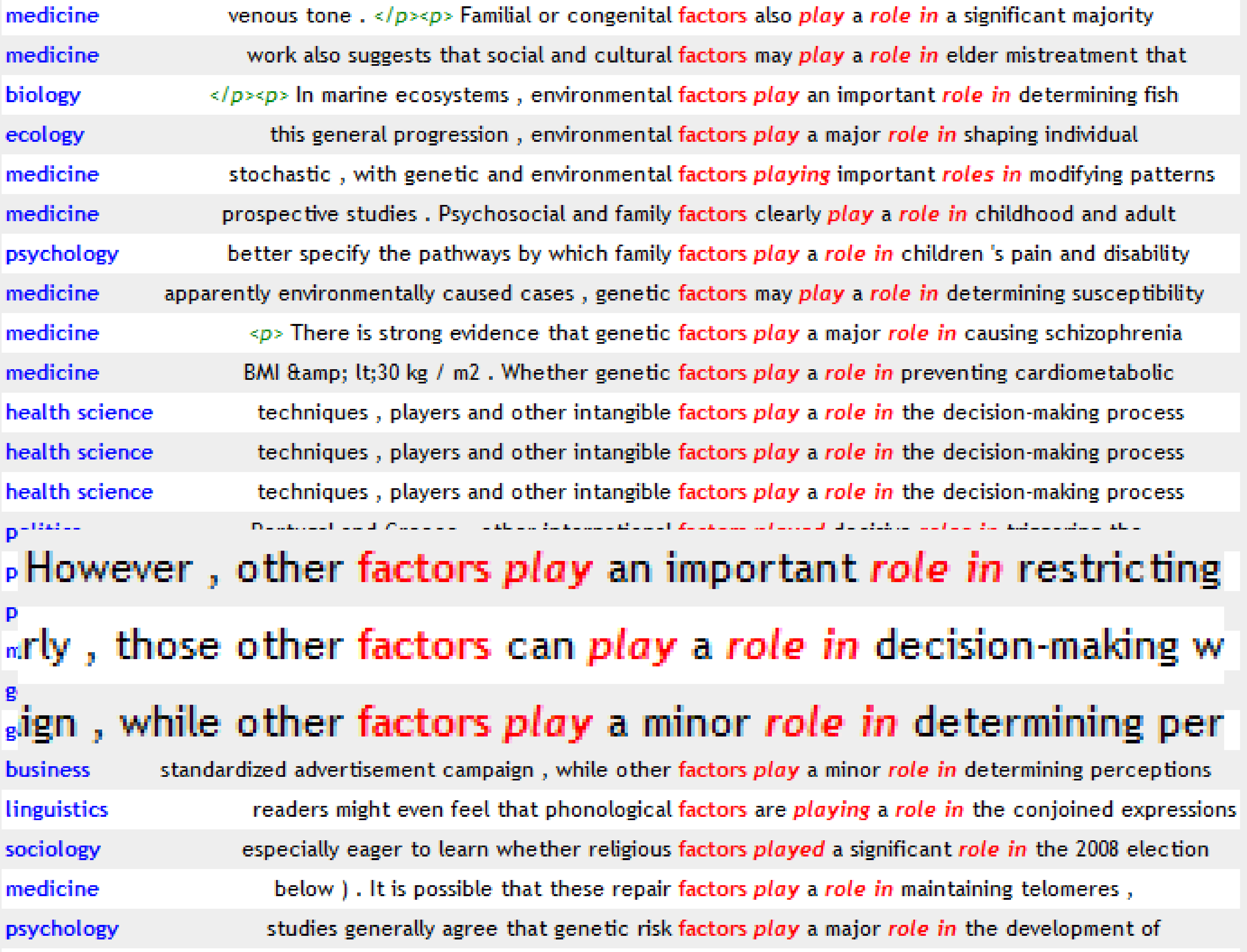
PREP N*	2445
of	645
in	314
with	264
to	247
on	190
for	177
than	108
as	91
from	77
by	73
between	58
according_to	55
at	33
into	23
around	16
below	9
since	6
without	6
about	6
under	[illegible]



maths	analysing the data in isolation to	*yield*	an	estimate	for the value of the parameter , it makes
maths	which of these two approaches	*yields*	an	estimate	closer to the true value . Unfortunately
maths	we knew that a method usually	*yielded*	an	estimate	which was very near to the true value of
education	and in those with Jean Dreze)	*yields*	an	estimate	of 44 million missing women in China ,
business	performance almost never	*yields*	empirical	estimates	that correspond to the conceptual model
ecology	other turtle species ,	*yielded*	an overall	estimate	of Ne over evolutionary time of between
ecology	Eight of the 10 loci	*yield*	very similar FST	estimates	between 0.055 and 0.072 . Because of this
earth sciences	the water flow (m2) of course	*yields*	an	estimate	of the discharge (m3 s-1) . *</p><p>* Figure
physics	calculation with the Standard Model	*yields*	an	estimate	of 1031 e cm , but other theories of '
business	outside view is more likely to	*yield*	realistic	estimates	, giving some protection against wildly
business	from such data limitations	*yields*	biased	estimates	of model coefficients . To ensure that
politics	This approach	*yields*	dramatically lower	estimates	of the impact of
economic					
biology	estimated by block jackknife) , this	*yielded*	an	estimate	of f^=14 % . However , shows that this
economics	is the double log , which	*yields*	direct	estimates	of elasticities but constrains the elasticity
economics	propensity score matching technique	*yields*	robust	estimates	of the ATE . *</p><p>* The results of the first-stage
sociology	patients under Open Access would	*yield*	an	estimate	of the magnitude of error in the data .
health science	. Such an approach should	*yield*	a better	estimate	than single equations ; it should also
biology	observed and interpolated cost data	*yields*	an	estimate	of global average variable cost per tonne
medicine	likelihood because it	*yields*	optimal parameter	estimates	with continuous multivariate normally distributed
earth sciences	Al-in-Opx thermometry	*yield*	extreme temperature	estimates	of c . 1010C for both samples . *</p><p>* P-T
medicine	least restrictive definition	*yielded*	an	estimate	of 11 % . When certainty criteria were
medicine	ethnographic and survey data) can	*yield*	reliable	estimates	of behavioural processes that can be scaled
biology	based on self-sampling can	*yield*	accurate	estimates	of catch and effort accounted for by the
ecology	without freezing does not	*yield*	numerical	estimates	, unless the method is properly calibrated

What makes a good dictionary example?

- *Taylor makes the following argument: …*
- *This approach yields dramatically lower estimates.*
- *Several other factors played a role in the decision-making.*
- *The most persuasive argument against this idea comes from Foster (2009).*



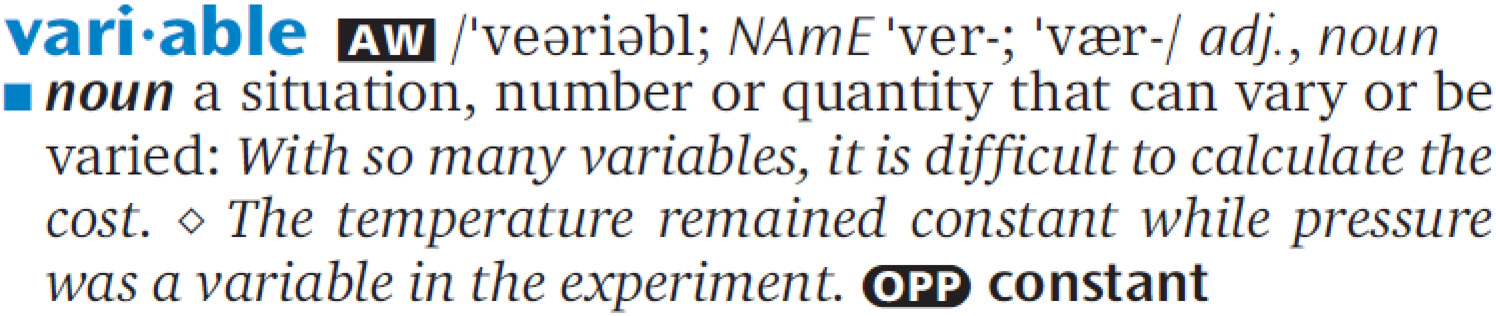
medicine	venous tone . </p><p> Familial or congenital factors also play a role in a significant majority
medicine	work also suggests that social and cultural factors may play a role in elder mistreatment that
biology	</p><p> In marine ecosystems , environmental factors play an important role in determining fish
ecology	this general progression , environmental factors play a major role in shaping individual
medicine	stochastic , with genetic and environmental factors playing important roles in modifying patterns
medicine	prospective studies . Psychosocial and family factors clearly play a role in childhood and adult
psychology	better specify the pathways by which family factors play a role in children 's pain and disability
medicine	apparently environmentally caused cases , genetic factors may play a role in determining susceptibility
medicine	<p> There is strong evidence that genetic factors play a major role in causing schizophrenia
medicine	BMI &amp; lt;30 kg / m2 . Whether genetic factors play a role in preventing cardiometabolic
health science	techniques , players and other intangible factors play a role in the decision-making process
health science	techniques , players and other intangible factors play a role in the decision-making process
health science	techniques , players and other intangible factors play a role in the decision-making process
politics	Portugal and Greece , other international factors played decisive roles in triggering the
p	However , other factors play an important role in restricting
p	
m	rly , those other factors can play a role in decision-making w
g	
g	ign , while other factors play a minor role in determining per
business	standardized advertisement campaign , while other factors play a minor role in determining perceptions
linguistics	readers might even feel that phonological factors are playing a role in the conjoined expressions
sociology	especially eager to learn whether religious factors played a significant role in the 2008 election
medicine	below) . It is possible that these repair factors play a role in maintaining telomeres ,
psychology	studies generally agree that genetic risk factors play a major role in the development of

variable *noun*

OXFORD
UNIVERSITY PRESS

Experiments are confined to a very narrow range of variables.

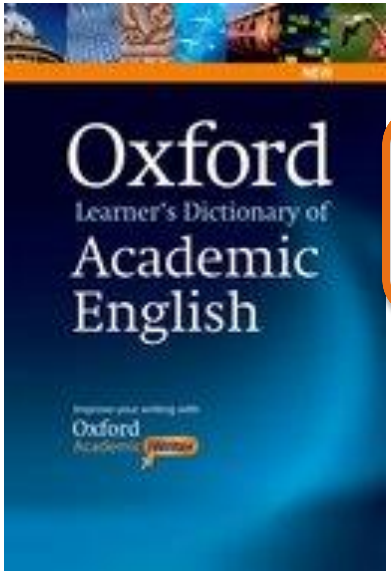
Drummond H. (2000) *An Introduction to Organizational Behaviour.* Oxford: Oxford University Press

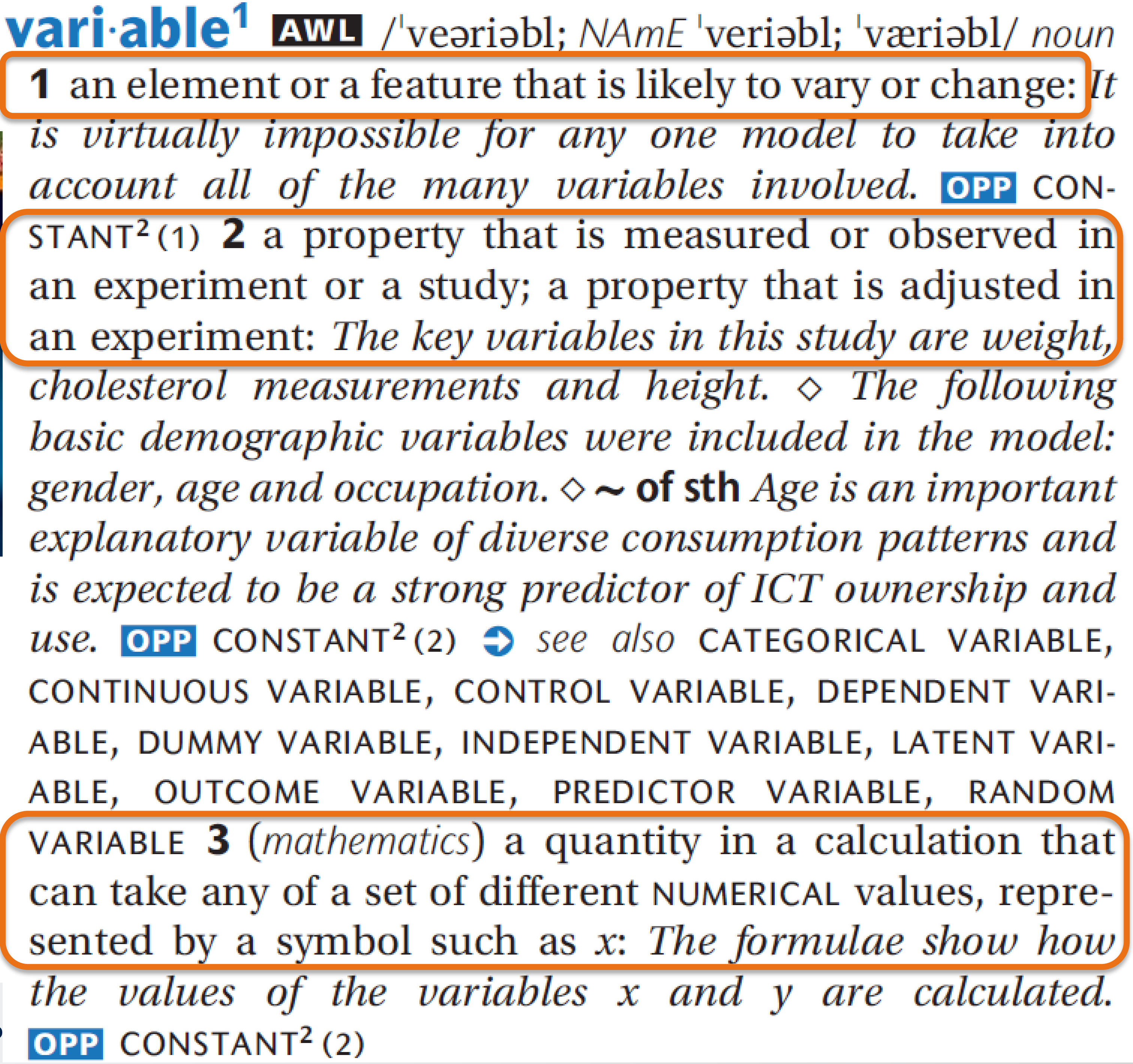


vari·able **AW** /'veəriəbl; *NAmE* 'ver-; 'vær-/ *adj.*, *noun*

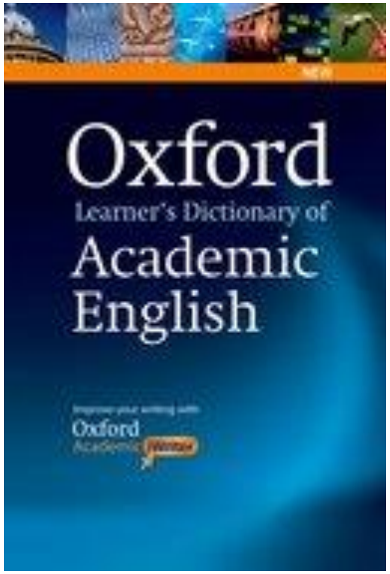
■ ***noun*** a situation, number or quantity that can vary or be varied: *With so many variables, it is difficult to calculate the cost.* ◇ *The temperature remained constant while pressure was a variable in the experiment.* **OPP** **constant**

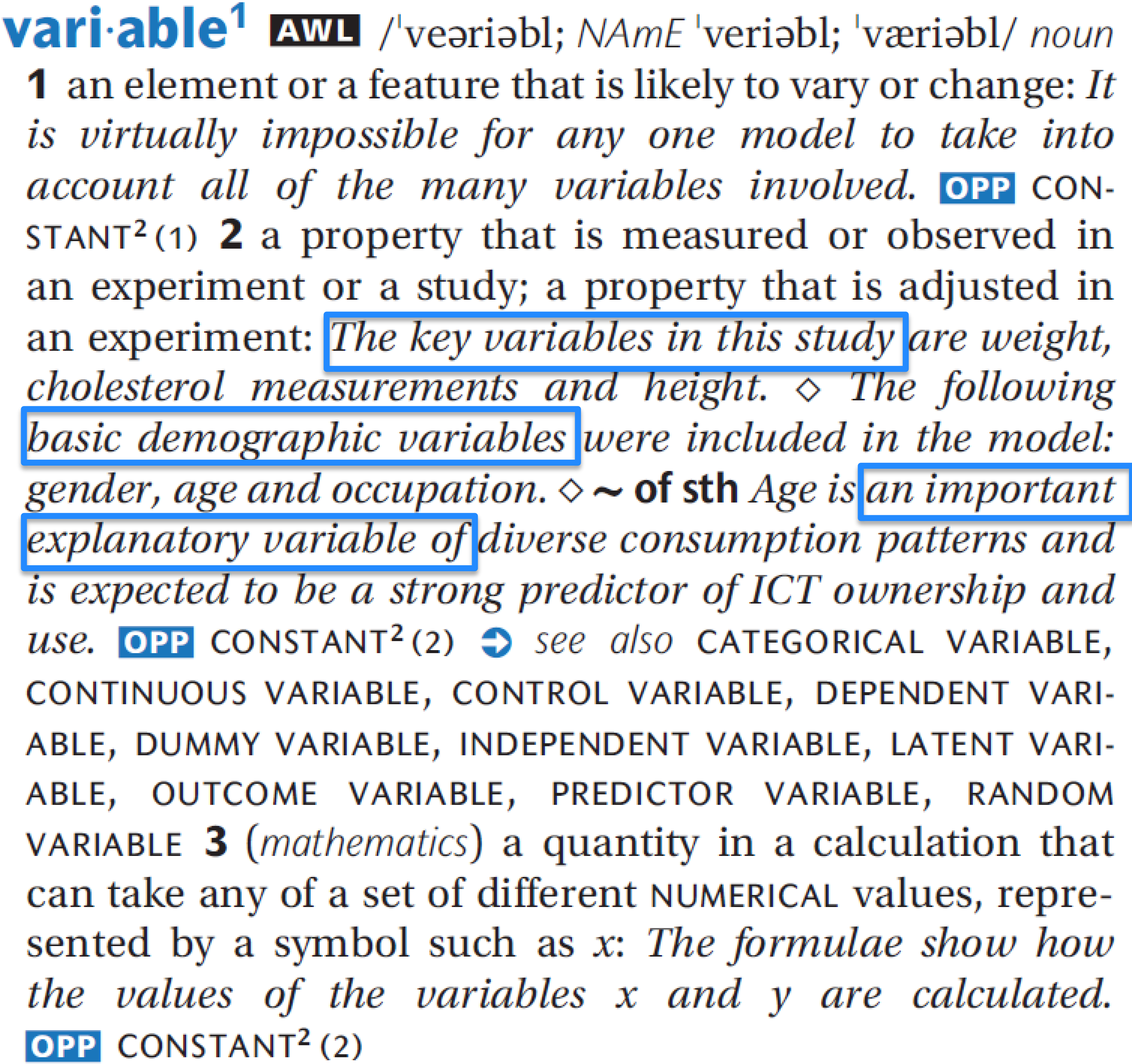
Oxford Advanced Learner's Dictionary, 8th edition (2010)





vari·able^{1} AWL /ˈveəriəbl; *NAmE* ˈveriəbl; ˈværiəbl/ *noun* **1** an element or a feature that is likely to vary or change: *It is virtually impossible for any one model to take into account all of the many variables involved.* OPP CONSTANT^{2} (1) **2** a property that is measured or observed in an experiment or a study; a property that is adjusted in an experiment: *The key variables in this study are weight, cholesterol measurements and height.* ◇ *The following basic demographic variables were included in the model: gender, age and occupation.* ◇ **~ of sth** *Age is an important explanatory variable of diverse consumption patterns and is expected to be a strong predictor of ICT ownership and use.* OPP CONSTANT^{2} (2) ➔ *see also* CATEGORICAL VARIABLE, CONTINUOUS VARIABLE, CONTROL VARIABLE, DEPENDENT VARIABLE, DUMMY VARIABLE, INDEPENDENT VARIABLE, LATENT VARIABLE, OUTCOME VARIABLE, PREDICTOR VARIABLE, RANDOM VARIABLE **3** (*mathematics*) a quantity in a calculation that can take any of a set of different NUMERICAL values, represented by a symbol such as *x*: *The formulae show how the values of the variables x and y are calculated.* OPP CONSTANT^{2} (2)





vari·able¹ AWL /ˈveəriəbl; *NAmE* ˈveriəbl; ˈværiəbl/ *noun* **1** an element or a feature that is likely to vary or change: *It is virtually impossible for any one model to take into account all of the many variables involved.* OPP CONSTANT² (1) **2** a property that is measured or observed in an experiment or a study; a property that is adjusted in an experiment: *The key variables in this study are weight, cholesterol measurements and height.* ◇ *The following basic demographic variables were included in the model: gender, age and occupation.* ◇ **~ of sth** *Age is an important explanatory variable of diverse consumption patterns and is expected to be a strong predictor of ICT ownership and use.* OPP CONSTANT² (2) ➔ *see also* CATEGORICAL VARIABLE, CONTINUOUS VARIABLE, CONTROL VARIABLE, DEPENDENT VARIABLE, DUMMY VARIABLE, INDEPENDENT VARIABLE, LATENT VARIABLE, OUTCOME VARIABLE, PREDICTOR VARIABLE, RANDOM VARIABLE **3** (*mathematics*) a quantity in a calculation that can take any of a set of different NUMERICAL values, represented by a symbol such as *x*: *The formulae show how the values of the variables x and y are calculated.* OPP CONSTANT² (2)

- Academic language is different.
- EAP students need a learner's dictionary based on analysis of real academic language.