



What do learners [^] actually *do* with corpus data?

Friday 7th June

14.15- 15.00

Tom Cobb

Université du Québec à Montréal

Get this PPT

at https://lextutor.ca/bawe_conf_tom.pptx (6.5 Mb)

Looking over old PPTS

- Presentations about computers in language learning used to be quasi-promotional
 - ‘CALL’
- Promotion is no longer needed
 - All over education, computer as learning tool is here to stay, for better or worse
 - Facebook’s “Summit” Saga
 - And in language learning...
 - *Rosetta, Babel, Duoling, One grain of rice, Memrise...*
 - Probably many current learners’ main experience of instructed learning

However...

- Many computational language learning tools have good record keeping & motivational/game elements but simplistic learning models
 - Mainly single-word learning by imitation and association
 - OK as far as it goes
 - Zero research
 - » (other than market research)
- Now what needs promotion is **what to do** with computers in learning
 - Through integration with SLA research

One CALL idea with some research behind it

- “Data driven language learning”
 - As a proven, research-based, computer-assisted component of a language learning program
- Meta-analysis on DDL in LL (2017)
 - Over 20 years, many uses, and varied environments...
 - Within-groups effect size (pre-post) effect size: $d = 1.5$
 - Between-groups e.s. (experimental-control) : $d = .95$
 - Both “strong for the field” (~Plonsky)
- And not exactly a specialty interest
 - *Language Learning*’s most-cited paper in 2017

EMPIRICAL STUDY



Corpus Use in Language Learning A Meta-Analysis

Alex Boulton and Tom Cobb

Université de Lorraine and Université du Québec à Montréal

This study applied systematic meta-analytic procedures to summarize findings from experimental and quasi-experimental investigations into the effectiveness of using the tools and techniques of corpus linguistics for second language learning or use, here referred to as data-driven learning (DDL). Analysis of 64 separate studies representing 88 unique samples reporting sufficient data indicated that DDL approaches result in large overall effects for both control/experimental group comparisons ($d = 0.95$) and for pre/posttest designs ($d = 1.50$). Further investigation of moderator variables revealed that small effect sizes were generally tied to small sample sizes. Research has barely begun in some key areas, and durability/transfer of learning through delayed posttesting remains an area in need of further investigation. Although DDL research demonstrably improved over the period investigated, further changes in practice and reporting are recommended.



TOP DOWNLOADED ARTICLE 2017-2018

CONGRATULATIONS TO

Thomas Cobb

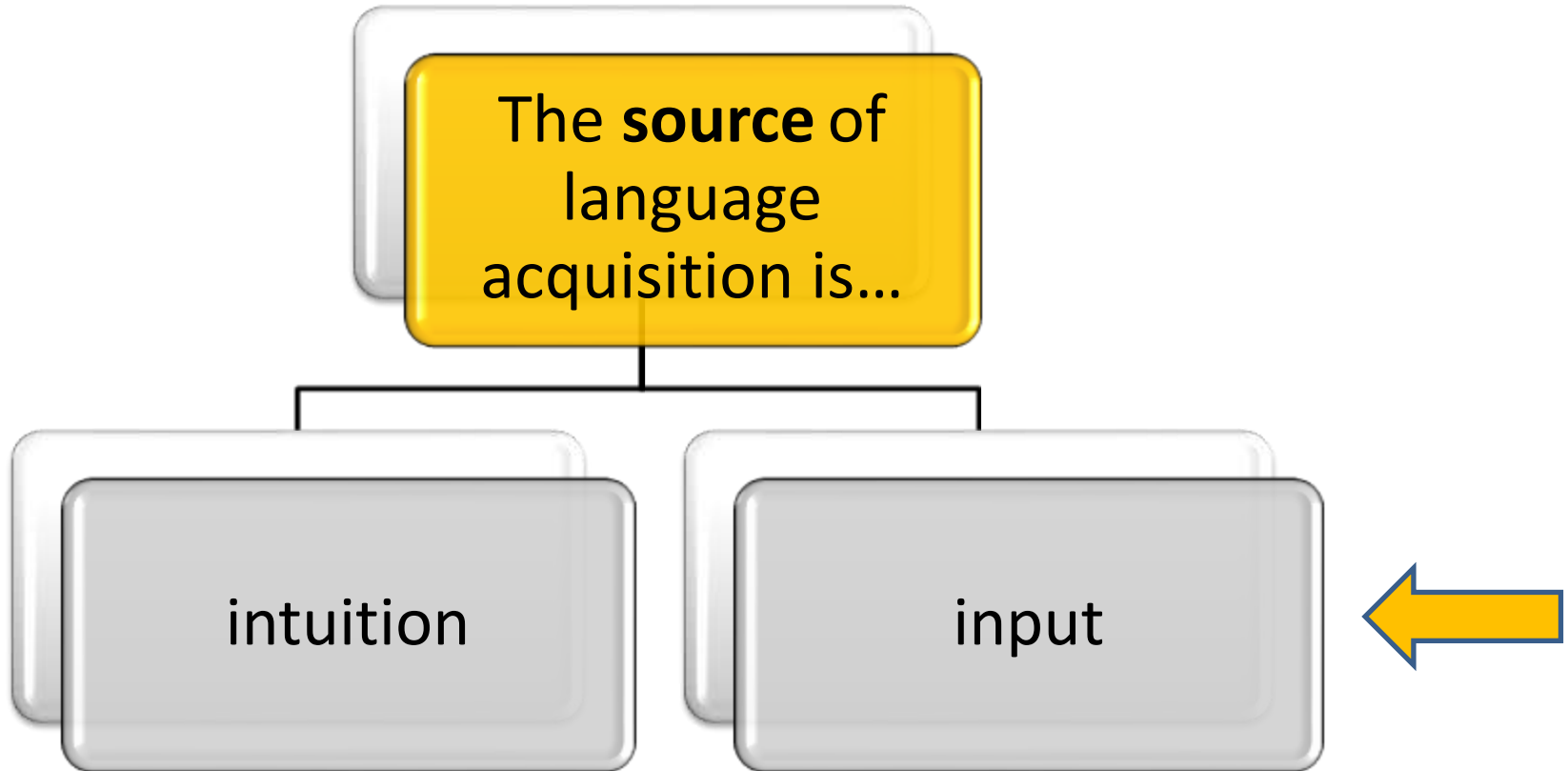
whose paper has been recognized as
a top 20 most read paper in

Language Learning

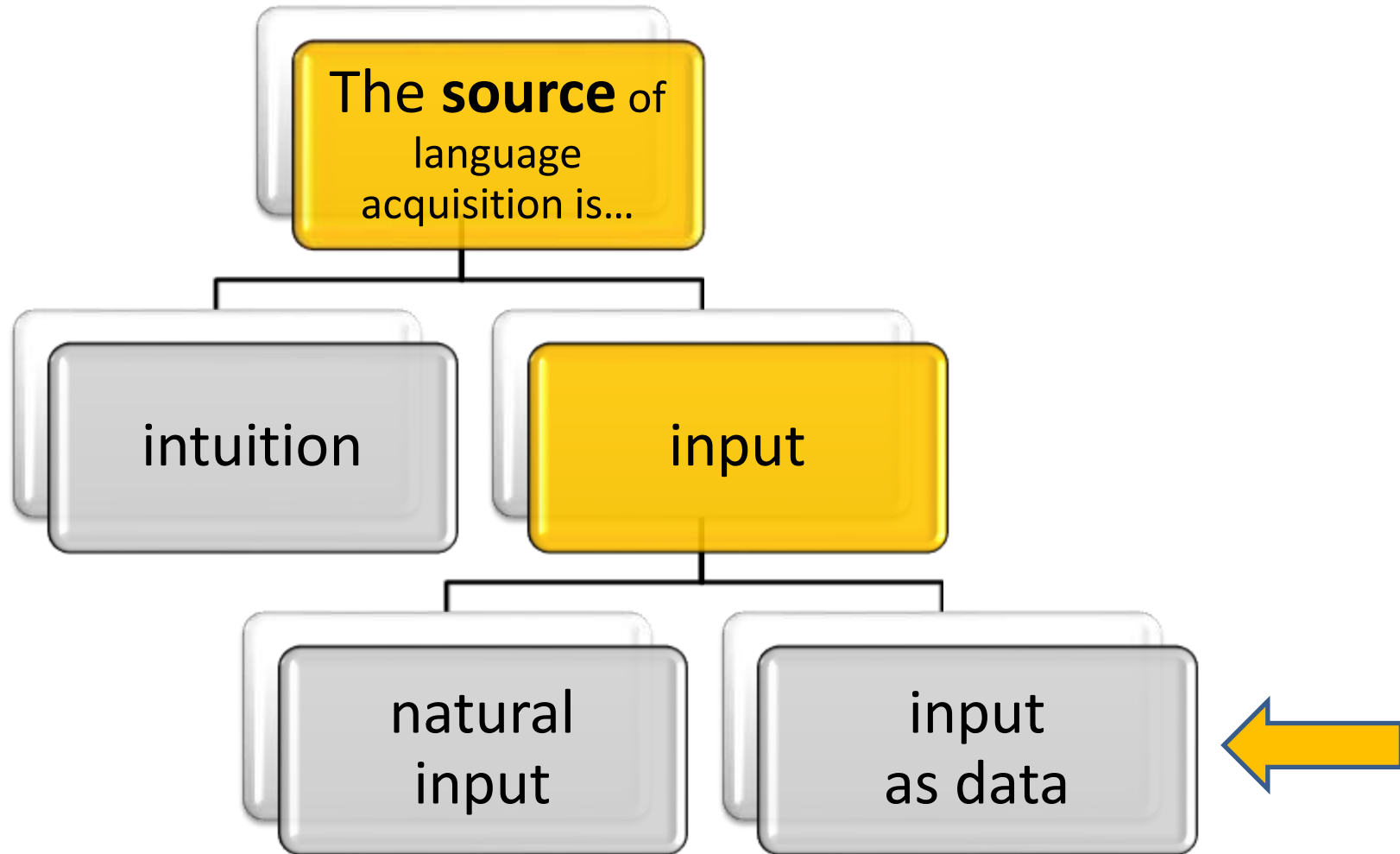
WILEY

What is Data-Driven Learning?

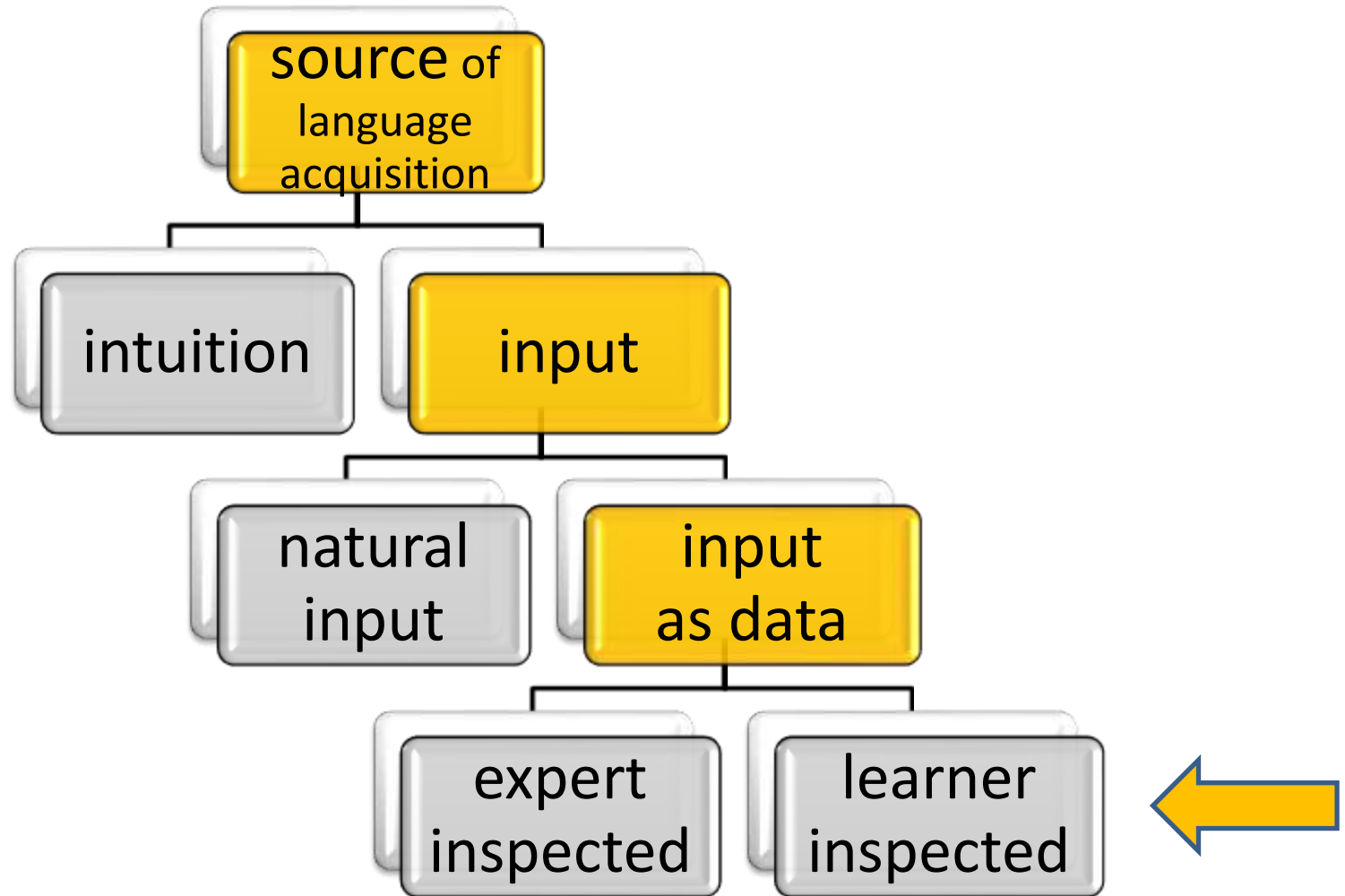
What questions does it still raise?



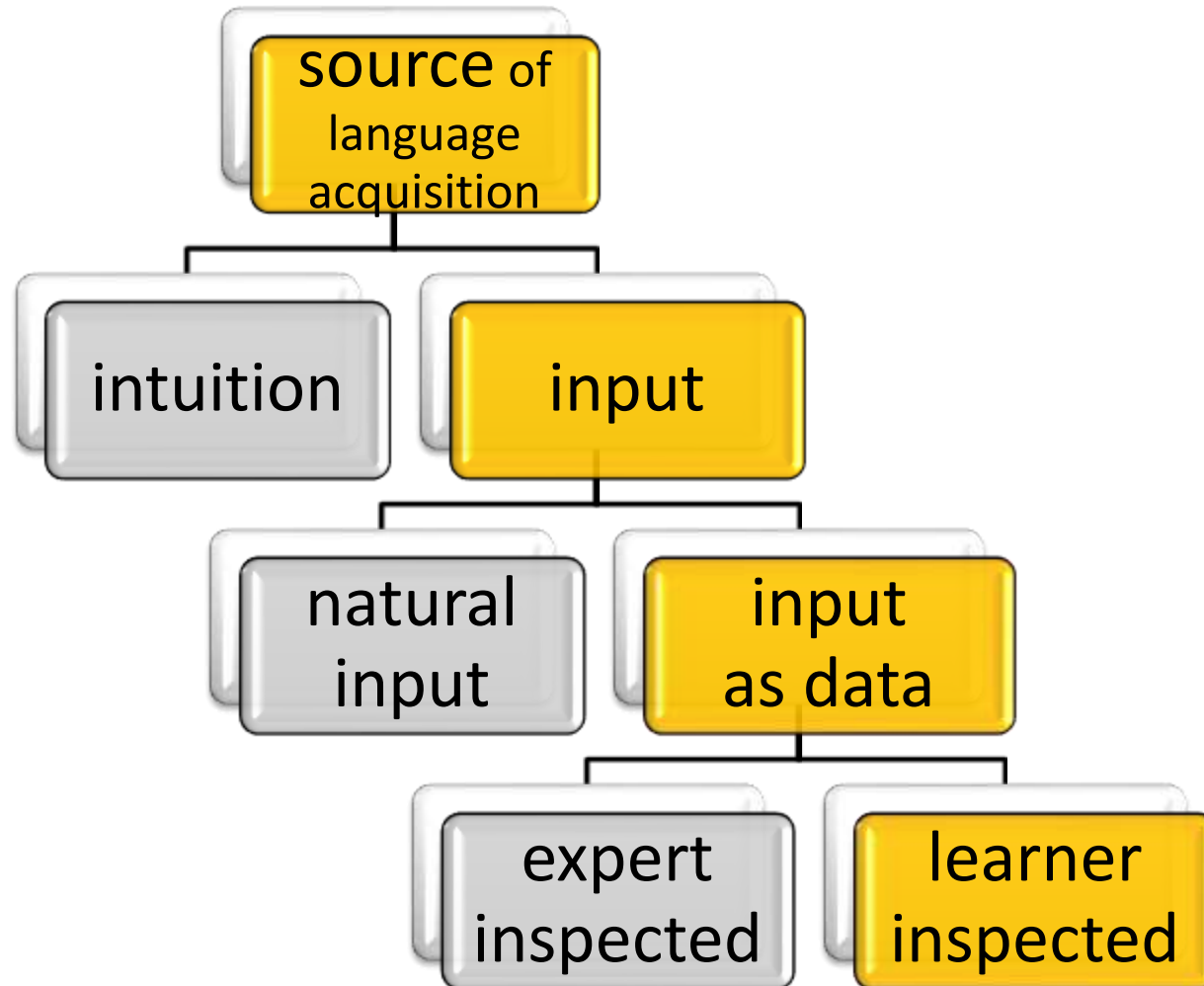
What is Data-Driven Learning?



What is Data-Driven Learning?



What is Data-Driven Learning?



What is Data-Driven Learning?

- Lextutor's *Web Concordancer* is an extended test of the 'learner inspected' model
 - A free resource, easy access, including mobile phone, used by about 5,000 users per week
 - Presumably teachers & their students
 - (more on this later)



8 Examples of “input as data,” or, learnability-enhanced input transformed by computer software

- Note :

- this means where the input is
significantly processed by software

- Not just Skype, e-pen-pals, Gutenberg or
Google books, online submission
 - Valuable as those may be

Enhancement 1:

Data assembly

- Problem: if a pattern is not clear from one piece of input...

the issue by granting an injunction in advance of publication." **If the plaintiff can prove immediately and convincingly that the defendant is intending to publish palpable untruths, an injunction could be granted.**

Otherwise, the rule against prior restraint must prevail in libel actions.

... it may be clear from several

attraction. Their movement consisted of magnetized bodies magnetizing one another." As for the fedayeen, they "observed a smiling rigour. **But the eroticism was palpable**. I could sense its vibrations, though I wasn't bothered by it" (p.126). But this eroticism, and the masculinity of these men, frequently confounds the

in present tariff charges upon the necessaries of life. **The benefits of such a reduction would be palpable and substantial, seen and felt by thousands who would be better fed and better clothed and better sheltered.** These gifts should be the willing benefactions of a Government whose highest function

Enhancement 2:

Useful effort-level raising

- Problem: Language work needs cognitive engagement to be effective
 - *Involvement load hypothesis*
 - Laufer & Hulsteyn (2001)
- Yet just raising effort *per se* is not efficient
 - Writing lines?
 - Committing texts to memory?
- Solution:
 - Corpus work raises effort usefully

Why usefully?

E.g. concordance vs. dictionary look-up for novel vocabulary

- Dictionary is an easy strategy but with limited involvement or transferability
 - To real life & work
- While assembling/examining/interpreting language data is a highly transferable strategy
 - Can become a habit of mind
 - **And produces stronger knowledge with better transfer to novel contexts**
 - **Cobb (1997)**

Enhancement 3 :

Data transformation to compensate for gaps in input

- Example: TTS

The screenshot displays the Lextutor interface for a TTS application. At the top, there are navigation links and a 'Save on Lextutor' button. The main content area shows a 'HYPERTEXT FILE: hors_de_prix_trump[1]' with a 'Dico' link and instructions for using the 'DEFINE' and 'SPEAK' buttons. Below this is a 'WORD BOX' containing the word 'estimée'. At the bottom, there are buttons for 'Dictator', 'MultiConc.', 'ID-Words', 'DEFINE', and 'SPEAK'. A 'Dictionary Window' is open on the right, showing a list of dates: 'MAY 2019', 'NOV 2018', 'APRIL 2013', 'SEPT 2013', 'JAN 2012', 'MARCH 2012', and 'DEC 2012'. A red arrow points from the 'SPEAK' button to the dictionary window.

With a fortune estimated at 3.5 billion dollars Donald Trump is the richest president in the history of the United States. Before becoming chief of state he multiplied the adventures financial and was able to enjoy a life of luxury which many would dream of.

Let's take a look at ten things Donald Trump owns! It's about the main residence of Trump his wife Melania and their sons Barron.

Spread over three floors the apartment is the image of luxury and excess. The rooms are covered in marble and diamonds without mentioning the frescoes on the

Dictionary Window

HYPERTEXT 2

TTS limit 100=>1200 chars. MAY 2019
iPad compliant NOV 2018
No-plug_in TTS APRIL 2013
Multi-Word TTS SEPT 2013
Wordreference JAN 2012
Cross browser MARCH 2012
Bilingual DEC 2012

Enhancement 3 : Data transformation to compensate for gaps in input

- Problem: Spoken words map to written words but the mapping is not obvious
 - Words recognized in writing are not recognized in speech, and v-v.
- Solution: TTS technology translates text to speech
 - makes every written word, sentence etc. click-pronounceable

Enhancement 4 :

Data transformation to focus on a research-indicated problem skill

Example: Text de/re-construction

The screenshot shows a Lextutor cloze test interface. At the top, the breadcrumb trail is "Home > Cloze > N-Word Input > Cloze for in_progress_20 | Ratio=7". A yellow box says "Save on Lextutor as David_Bowie_-_Major_Tom[3].html". A red control bar shows "Qs: 20", "Correct: 0", "Tries: 0", and a "CHECK" button. Below this is a "Do again" button and a "History >>" link. The main content area features a video player showing a profile of Major Tom in a space helmet. To the right, a list of words is displayed in a vertical column: "on", "helmet", "leave", "much", "the", "think", "dare", "door", "peculiar", "today", "can", "blue", "do", "miles", "go", "knows", "she", "knows", "circuit", "is", "dead", "wrong", "hear", "me", "Major Tom?". The text of the cloze test is displayed on the right, with numbered blanks corresponding to the word list. The text includes: "Ground control to Major Tom, Ground control to Major Tom, Take your protein pills and put your [1] on Ground control to Major Tom, Commencing countdown, engines on Check ignition and may God's [2] you This is ground control [3] Major You've really made the grade And the papers want to know whose [4] you wear Now it is time to [5] the capsule [6] dare Stepping through [6] door And I am floating in a [7] peculiar way And the stars look very [8] today Here am I sitting in a [9] can far above the world Planet Earth [10] blue and there is nothing I [11] do Though I am past one hundred [12] miles, I am feeling very still And I [13] my spaceship knows which way to go Tell my wife I love her very [14], she knows Ground control to Major Tom, [15] circuit is dead, there is something wrong Can you hear me, Major Tom? Can you hear me, Major Tom?" At the bottom, a red box contains the text "ANSWERS: Click for Eng_Eng definitions" and a list of words: "can (2) different hear helmet is (2) leave love most much my shirts the think thousand tin to you your".

Enhancement 4

Data transformation to focus on a skill

Example: Text de/re-construction

- Word knowledge is consolidated through many retrievals

Problem: Opportunities for word-to-meaning retrieval are numerous

– ... while meaning-to-word retrieval is rare

- Productive lexicon = 40% of receptive

- Solution: Transform written/spoken texts into cloze passage

- Retrieve word/phrase from overall meaning
- Is effortless for teacher/computer to make
- Is usefully effortful for learner

+ Clearly these learnability enhancements can function in consort

– Combining data-assembly & meaning-word retrieval

MultiConc Output

[Printing this up?](#)

Multiple Concordances (Corpus=*bawe_sampler.txt*)

| HOWEVER | FOR EXAMPLE | FURTHERMORE | IN OTHER WORDS

Hits: 007

[001] could mature, as the family breaks up.	_____ , it is the "traditional authoritarian str
[002] ecause of the appalling economic climate.	_____ , this is not as effective as the techniqu
[003] ms of revolutionary cinematic techniques.	_____ , unlike in Triumph des Willens, Kuhle Wa
[004] ade was it released for general showing."	_____ , it seems that Brecht asks quite a lot of
[005] ns she reads about in her beloved novels.	_____ , this essay will also attempt to establis
[006] se of what "real" life and love are like.	_____ , Emma quickly realises that the novels ar
[007] h and pray before her first visit to him.	_____ , she gives into temptation and, as a marr

Hits: 007

[001] structure and activity in related GPCRs,	_____ the above mentioned olfactory GPCRs. The
[002] rrange of seven-transmembrane helices	_____ is stabilised by molecular interactions,
[003] tions and with high specificity - it can,	_____ , distinguish between the R- and S-isomer
[004] ons: recognising and binding to antigens,	_____ a bacterial toxin in order to inhibit its
[005] a rational justification for the policy,	_____ sending out a strict message of no-tolera
[006] t and Y's relationship with this teacher.	_____ , how long ago was the complaint made? How
[007] eed to know more about Y's circumstances,	_____ whether she was a dealer. However, if the

Hits: 007

[001] right to security and freedom of speech.	_____ , the word guerre can be appropriate given
[002] 2+ ions, which converts cAMP back to AMP.	_____ cAMP itself activates protein kinase A, w
[003] th respect to their biological functions.	_____ , one of the most characteristic structura

Enhancement 5 :

Data-Driven Instructional Design

E.g., Assess difficulty of text through frequency profiling

Problem: Teachers and materials writers have a poor sense of word frequency

Or what constitutes an appropriate text for Ss at a certain level

Solution: Select or modify texts according to a frequency profile **(VP)**

Enhancement 5

2 texts - Can native speaker discern which is 'hard' ?

A: Well, it was a narrow escape. But we did it.

Canadians have preserved their liberties and
independence
American be

We knew
the United States
genueflect to
dragging God
remorselessly
since 9/11.

They have
simpleton's g
banishing nu
invective pos
passing itself

It's a diffi
United States
in all the dus
But we did it

Profile summary		
K	#tokens	cumul%
K-1	92	72.4
K-2	9	79.5
K-3	7	85.0
K-4	3	87.4
K-6	2	89.0
K-7	4	92.1
K-10	1	92.9
K-12	2	94.5
K-15	2	96.1
K-16	1	96.9
OFF	4	≈100

wds

B: My name is Dr Frederick Treves. I am a doctor
at the London Hospital

One day
of a shop
the shop
intereste

It was a
in the pic
me. He did not look like a man. He looked like an
elephant. I read the writing under the picture. It
said: Come in and see the Elephant Man. 2 pence.

I opened the door and went in. There was a
man in the shop. He was a dirty man in an old
coat with a cigarette in his mouth. 'What do you
want?' he asked. 'I'd like to see the elephant man,
please,' I said. The man looked at me angrily.
'Well, you can't,' he said. 'The shop's closing now.
You can come back tomorrow.'

Profile summary		
K	#tokens	cumul%
K-1	169	97.7
K-2	4	100.0
OFF	0	≈100

172 wds 22

Enhancement 6 :

Data assembly again, but *(a)* by learners & *(b)* with a network

Example

- Problem:
 - Too many new words for learners to look up
 - In reading a challenging text
- Solution:
 - Divide the look-ups
 - then re-assemble with a network



Group Lex M

Group Lex M

«demo_lab_r

QUIZ 1 - contexts from WHOLE GP LEX 20 May 19, 9:07

Quiz 2 - new contexts - after 100% on Quiz 1

Check Questions: 10 Correct: 0 Tries: 0 Percent: 0

History >>

WORD	Example
1 perennial	The debate on whether voice is teachable in L2 classroom is perennial .
2 insight	Endurance could provide insight into the environmental history of the art but there are risks.
3 betterment	I believe politics politicians are vit the betterment society.
4 dovetail	Listening and communication skills must dovetail with other skills.
5 cadre	- He trained cadre engineers and business schools in Iraq.
6 soluble	Not all problems soluble , not all risks can be hedged at acceptable cost.
7 hypnotize	I was fascinated them, hypnotize them - until you wake up and realise they're nuts.
8 dogma	On the contrary, renew your doctrines and dogmas , make them relevant to times in which you seek support.
9 contiguity	Contiguity is necessary in all forms of learning.

	NEW WORD	EXAMPLE	WORD CLASS	DEFINITION
1		I believe politics and politicians are vital to the betterment of society.	N	The improvement of something.
2		Not all problems are <input type="text"/> , not all risks <input type="text"/> ed at acceptable cost.	ADJ	(of a problem) able to be solved.
3		On the contrary, you renew your doctrines and betterment s, making them relevant to the times in which you seek support.	N	A principle or set of principles laid down as incontrovertibly true.
4		Endurance could provide <input type="text"/> into the cadre al history of the area, but there are risks.	N	The capacity to gain an accurate and understanding of someone or something.
5		Listening and communication skills must <input type="text"/> other skills.	V	Fit or cause to fit together easily and interlock.
6		- He trained <input type="text"/> s of engineers and business schools in Iraq.	N	A small group of people specially trained for a purpose or profession.
7		Its elegant script is unknown from any other source to its <input type="text"/> content has emerged.	ADJ	Relating to meaning in language or logic.
8		I was fascinated by them, <input type="text"/> d by the insight wake up and realise they're nuts.	V	Capture the whole attention of (someone).
9		<input type="text"/> is necessary in all forms of learning.	N	The sequential occurrence or proximity of response, causing their association in learning.
10		The debate on whether voice is teachable in L2 classroom is <input type="text"/> .	ADJ	long-lasting

Enhancement 7 :

Automated transfer

So we have ~

- Data assembly
- Useful effort raising
- Contextual TTS
- Meaning-to-word retrieval
- Collaborative learning through a network...

+ the elusive *learning transfer* ?

Automated transfer

- Problem: Language learning is contextual and learning may not be transferred to a novel context
 - E.g., word knowledge
 - Word learned in one text
 - Is not recognized in another text
 - It is hard to get practice in transfer
 - Word is forgotten when re-encountered
- A DDL approach, again with Group Lex
 - Again ‘bringing it all together’ →

ALL ENTRIES:

Quiz Options>> [Print | Interactive] 1st 10 Any 20 None [+extras] **Quiz checked words**

#	Qz	WORD	EXAMPLE	POS	DEFINITION	GROUP	STU
1	<input checked="" type="checkbox"/>	perennial	The debate on whether voice is teachable in L2 classroom is perennial .	ADJ	long-lasting	Arts	John
2	<input checked="" type="checkbox"/>	insight	Endurance could provide insight into the environmental history of the area, but there are risks.	N	The capacity to gain an accurate and deep understanding of someone or something.	Arts	Muh
3	<input checked="" type="checkbox"/>	betterment	I believe politics and politicians are vital to the betterment of society.	N	The improvement of something.	Arts	Muh
4	<input checked="" type="checkbox"/>	dovetail	Listening and communication skills must dovetail with other skills.	V	Fit or cause to fit together easily and conveniently.	Arts	Muh
5	<input checked="" type="checkbox"/>	cadre	- He trained cadres of engineers and built health clinics and schools in Iraq.	N	A small group of people specially trained for a particular purpose or profession.	Arts	Muh
6	<input checked="" type="checkbox"/>	soluble	Not all problems are soluble , not all risks can be hedged at acceptable cost.	ADJ	(of a problem) able to be solved.	Arts	Muh
7	<input checked="" type="checkbox"/>	hypnotize	I was fascinated by them, hypnotized by them - until you wake up and realise they're nuts.	V	Capture the whole attention of (someone); fascinate.	Arts	Muh
8	<input checked="" type="checkbox"/>	dogma	On the contrary, you renew your doctrines and dogmas , making them relevant to the times in which you seek support.	N	A principle or set of principles laid down by an authority as incontrovertibly true.	Arts	Muh
9	<input checked="" type="checkbox"/>	contiguity	Contiguity is necessary in all forms of learning	N	The sequential occurrence or proximity of stimulus and response, causing their association in the mind.	Arts	Muh
10	<input checked="" type="checkbox"/>	semantic	Its elegant script is unknown from any other source, and not one clue as to its semantic content has emerged.	ADJ	Relating to meaning in language or logic.	Arts	Muh
11	<input type="checkbox"/>	traverse	The reader traverse the beam in the opposite direction to that of the object under test to reduce image blur.	V	Move back and forth or sideways.	Arts	Muh
12	<input type="checkbox"/>	expound	'He used it as an opportunity to expound the idea of flexible	V	Present and explain (a theory or idea) in detail.	Arts	Muh

V.Mobile
on pad/phone

COMPUTER
See All 222...

Add Entry

Add Student

Edit Entries

Need Gp Lex?
Ask Tom

X-tract-list

Research?

How to GL

Speak the
selection

Start + click
NO SILENCE

* MODULAR! *

GROUP LEX v.8.2

- v. Road test
- + Dev Lab
- + Demo

OS=Win
BROWSER=Chr

V.Mobile
on
pad/phone

COMPUTER
See All 222...

Add Entry

Add Student

Edit Entries

Need Gp Lex?
Ask Tom

X-tract-list

Research?

How to GL

Speak the
selection

Select + click
No size limit

	NEW WORD	EXAMPLE	WORD CLASS	DEFINITION
1		I believe politics and politicians are vital to the betterment of society.	N	The improvement of something.
2		Not all problems are <input type="text"/> , not all risks <input type="text"/> ed at acceptable cost.	ADJ	(of a problem) able to be solved.
3		On the contrary, you renew your doctrines and betterment them relevant to the times in which you seek s	N	A principle or set of principles laid down as incontrovertibly true.
4		Endurance could provide <input type="text"/> into the cadre the area, but there are risks.	N	The capacity to gain an accurate and understanding of someone or someth
5		Listening and communication skills must <input type="text"/> contiguity other skills.	V	Fit or cause to fit together easily and
6		- He trained <input type="text"/> s of engineers and bu dogma schools in Iraq.	N	A small group of people specially train purpose or profession.
7		Its elegant script is unknown from any other so dovetail to its <input type="text"/> content has emerged.	ADJ	Relating to meaning in language or lo
8		I was fascinated by them, <input type="text"/> d by the hypnotize realise they're nuts.	V	Capture the whole attention of (someo
9		<input type="text"/> is necessary in all forms of learning insight	N	The sequential occurrence or proximi response, causing their association in
10		The debate on whether voice is teachable in L perennial <input type="text"/> .	ADJ	long-lasting

semantic

soluble

QUIZ 2 - Contexts from a corpus - >>
Choose corpus BAWE Brit.Academic (1+m) ▼ and then Go Quiz 2

	NEW WORD	EXAMPLE	WORD CLASS	DEFINITION
1		I believe politics and politicians are vital to the betterment of society.	N	The improvement of something.
2		Not all problems are soluble , not all risks can be hedged at acceptable cost.	ADJ	(of a problem) able to be solved.
3		On the contrary, you renew your doctrines and dogma s, making them relevant to the times in which you seek support.	N	A principle or set of principles laid down by an authority as incontrovertibly true.
4		Endurance could provide insight into the environmental history of the area, but there are risks.	N	The capacity to gain an accurate and deep understanding of someone or something.
5		Listening and communication skills must dovetail with other skills.	V	Fit or cause to fit together easily and conveniently.
6		- He trained cadre s of engineers and built health clinics and schools in Iraq.	N	A small group of people specially trained for a particular purpose or profession.
7		Its elegant script is unknown from any other source, and not one clue as to its semantic content has emerged.	ADJ	Relating to meaning in language or logic.
8		I was fascinated by them, hypnotize d by them - until you wake up and realise they're nuts.	V	Capture the whole attention of (someone); fascinate.
9		contiguity is necessary in all forms of learning	N	The sequential occurrence or proximity of stimulus and response, causing their association in the mind.
10		The debate on whether voice is teachable in L10 classroom is perennial .	ADJ	long-lasting

QUIZ 2 - Contexts from a corpus - >>
Choose corpus BAWE Brit.Academic (1+m) ▼ and then Go Quiz 2

QUIZ 2 - Contexts from a CORPUS (*bawe_sampler.txt*)

Which word/phrase fits all the gaps in each set? (Corpus=*bawe_sampler.txt*)

[betterment](#) [insight](#) [perennial](#) [semantic](#) [soluble](#)

Click words for Dictionary [Eng_Eng](#)

"DOVETAIL, CADRE, HYPNOTIZE, DOGMA, CONTIGUITY," not found in Corpus *bawe_sampler.txt*

Questions : 5 Completed: 0 Tries : 0 Per cent: 0 [History >>](#)

1.

[001] rfactants that contain a hydrophobic portion, which is _____ in oil-like solutions, and a hydrophilic
[002] il-like solutions, and a hydrophilic portion, which is _____ in water. This characteristic results
[003] to correspond with whether the plant stores starch or _____ sugars (1) and the precursor of these
[004] f n ? 5 then the general polynomial of degree n is not _____ by radicals. (The theorem is quoted fr
[005] ad to a better understanding. Section 6 We now define _____ groups, which Galois studied when work
[006] ynomials by radical. Definition 6.1 A finite group is _____ if it has a series of subgroups FORMU
[007] rphism theorems which will enhance our knowledge about _____ groups. Theorem 6.2 Suppose G is a gro
[008] level. We now look at some conditions for groups to be _____. This will be a great help in following

2.

[001] t _____ 08 Abstract: The organisation of the _____ mental lexicon is still an open questi
[002] e _____ duction The mental organisation of the _____ lexicon is an important question in co
[003] t BETTERMENT _____ in the environment that organises the _____ space. According to the first view (he
[004] f _____ ects share, the closer they are in the _____ space. For example, 'apple' and 'plum'
[005] d 'plum' are closer neighbours in the _____ space than 'apple' and 'basket'. As a
[006] d _____ ut not of 'basket'. The second view on _____ organisation (henceforth association-b
[007] ner should be closer neighbours in the _____ space than the latter. Indeed, when asl
[008] provide support for both theories of _____ organisation. Buchananal. (1999) showe

SEMANTIC

SOLUBLE

Enhancement 8 :

Negative evidence

Steven Pinker ~

The strongest argument for intuition as a source of language acquisition is the lack of ‘negative evidence’ in natural input... *(paraphrase)*

*How would a learner ever know what is ‘not language’?
(= not possible in his/her particular target language)*

A concordance assembly from a large corpus can show what does not exist in a language

For example, the Arab learner’s “I go to home” →

32 hits

Standardized to 5 per million (hits/corpus size x 1,000,000)

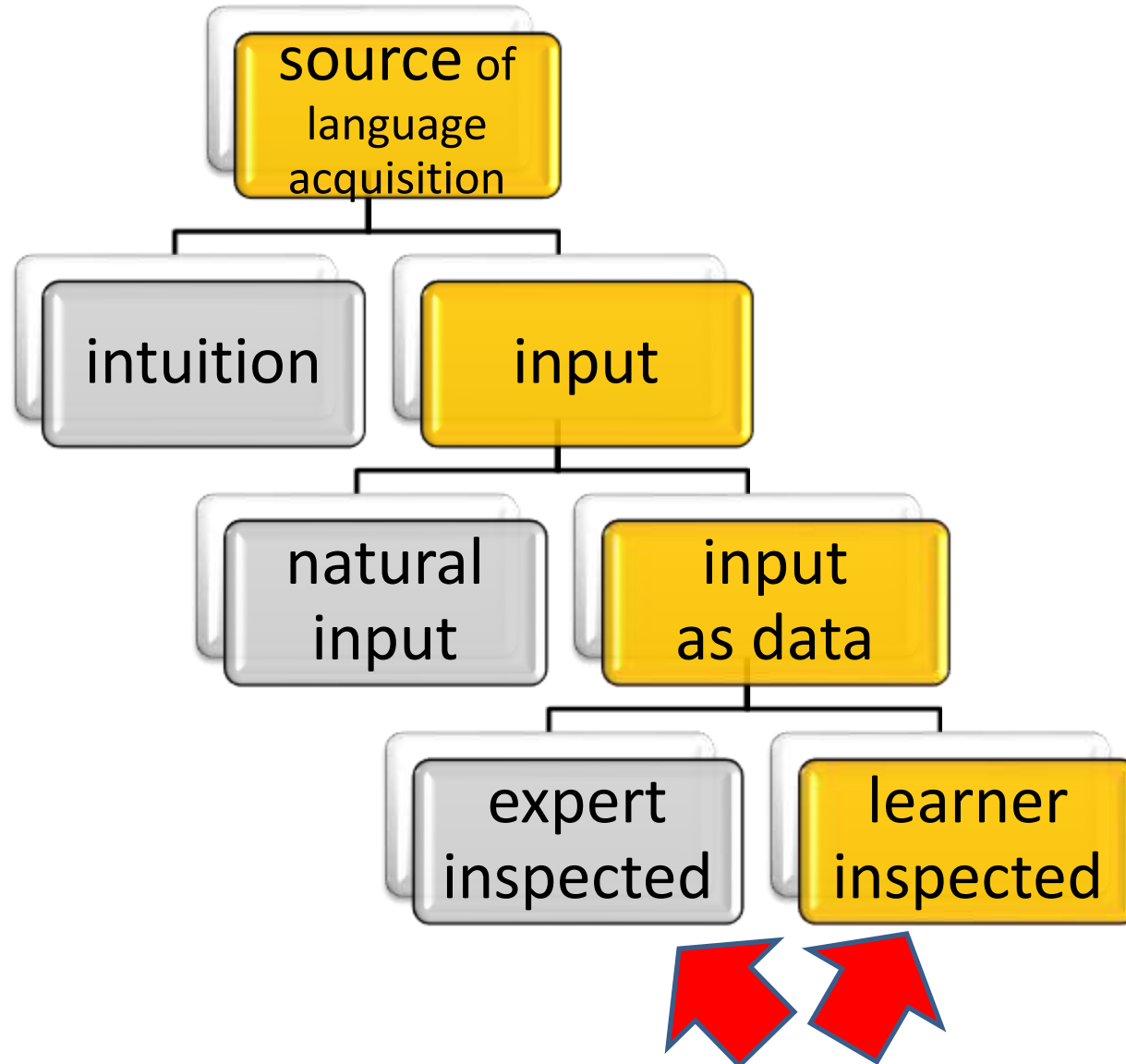
FREQUENCY BY MEMBER OR STEM: go=26 going=5 went=1

Click any [KEYWORD](#) for more context

- 001. can monitor you but you will be able to [GO home](#) once we have your current sympto [medicine]
- 002. - remained an upbeat man and wanted to [GO home](#). Social - had social support fro [medicine]
- 003. licking the browser back button. We can [GO to main page \(home button\)](#) but we can [computer_sci]
- 004. almost angry desire and desperation to [GO home](#) due to the onomatopoeic nature o [english]
- 005. e of care for you when you are ready to [GO home](#) that will help yourself and your [medicine]
- 006. tion for workers is the money that they [GO home](#) with at the end of the day. This [hospitality]
- 007. electrolyte imbalance. As Mr wishes to [GO home](#) to die, at present he cannot cop [medicine]
- 008. be arranged Every effort to allow Mr to [GO home](#) as he is very distressed in hosp [medicine]
- 009. es in Japan. The subsidiaries's profits [GO to the home country](#). The U.S. subsidi [business]
- 010. e prosthetic valve. ICE - Would like to [GO home](#) very soon and resume normal acti [medicine]
- 011. ically likely that they will be able to [GO home](#). Ensure that they are aware of h [medicine]
- 012. days. Some nights Mrs Stacy was able to [GO home](#) and care for Sam and get a good [health]
- 013. hen bourgeois women were not allowed to [GO](#) out of their [homes](#) as this represente [hospitality]
- 014. tion. An example including a command (""[GO home!](#)" is true iff the speaker ordere [philosophy]
- 015. " is true iff the speaker ordered me to [GO home](#)."") also seems to work as the int [philosophy]
- 016. s reported that some men shouted "Women [GO home!](#)" However, he suggested that for [us_studies]
- 017. e some idea of when he might be able to [GO home](#). He is also concerned about how [medicine]
- 018. uring university holidays when students [GO home](#). This results in the business su [hospitality]
- 019. major decrease in profit when students [GO home](#). In order to see whether promoti [hospitality]
- 020. be and is keen to be treated so she can [GO home](#). Management Initial investigatio [medicine]
- 021. acter, who constantly helps Odysseus to [GO home](#) and to settle all the things whi [classics]
- 022. Phaeacians who helped good Odysseus to [GO back home](#) safely, there is an obvious [classics]
- 023. er et Al, 2004). Freda does not want to [GO](#) into a nursing [home](#) because she feels [health]
- 024. weeks post-discharge. She was happy to [GO home](#) and reassured as to the explanat [medicine]
- 025. rated on EEG recording. He was happy to [GO home](#) and discharged on lamotrigine. I [medicine]

What is Data-Driven Learning ?

What questions does it still raise ?



Is it better...

- For learners to work with raw data themselves
 - Pushing the software in their own way to raise data comprehensibility & learnability
 - From a few hints and demos
- Or for teachers, course designers, materials writers
 - To do this *on learners' behalf*
 - Offering learners themselves just the final step in the process ?

Ex: Ss determine the collocations of *ride* and *drive* in a concordance that was basically set up to make this obvious

With messy or ambiguous data eliminated

A case study: Concordancing on Lextutor with BAWE

Home > [Concordancers](#) > Corpus search input (Eng)

Corpus Concordance English v.8

» With sub-corpus sort on * asterisked * corpora +NEW* BAWE Brit Academic Written (8 million/30 subs) » DEMO + Spoken (1.6m/9 subs) » DEMO

[French](#) [German](#) [Spanish](#) English Basic Speed \cong 1 second per million words of corpus

Add more for extras (associated words, family search, sub-corpus)

Keyword(s): (Max chars.) In corpus:

<< **Corpus descriptions**

- BAWE on Lextutor since 2018
 - >>>Well promoted
- One of about 25 open choice corpora & text collections

Corpus Concordance English

»With sub-corpus sort on * asterisked* corpora +NEW* BAWE Brit Academ
(8 million/30 subs) » DEMO + Spoken (1.6m/9 subs) » DEMO

[French](#) [German](#) [Spanish](#) English

Base Speed \cong 1 second per million
words of corpus

Add more for extras (associated words, family
search, sub-corpus)

Keyword(s): (Max chars.) In corp

<< Corpus descriptions

OPTION : With as se
words to sd
and NC
the line

CONTROLS :

Sorted By word
Gapped?
On STARTS or FAMIL

+ Scan for any recu

DEMOS : Demo 1
'Royal' in US v UK

Online Concordance cor

BNC Written (1 million)
BNC Spoken (1 million)
Brown_BNC Writ. (2+m)
BAWE Sampler (1+m)
BASE Spoken All (1.6m)*
BAWE Written All (8m)*
Brown_v BNC Writ. (2+m)*
Wiki Corpus (1+m)*
Brown Subs (520k)*
BNC Sp_v_Wr (2m)*
Graded readers (1.25m)*
Disney Scripts (104k)*

W SORT ↓ Sort Lines by Me
yword | Line Width
ion of forms before concordance lin
within words presenting

Demo 5 **sub-corpus >**
River bank **Money bank**
ytechnic; v.8 PERL code by

Research Questions

- Which resources do users exploit most in Lextutor's concordance program?
- Which resources do they exploit least?
- Which resources do they fail to notice?
- And is this same or different for BAWE/BASE corpora ?

What can the BAWE/BASE do? (as realized on Lextutor)

Main affordances

- **Size**
- **Speed**
 - < 10 secs for basic word + complete output ('sex'=2119 lines)
 - Cf Mark Davies' *Coca*, which gets speed by giving only 100 lines per screen
 - So everything cannot be seen together
- **Number and nature of academic sub-divisions**
- **By far Lextutor's best corpus**

Specific examples →

1. Count + examples by discipline

Re-Conc family sex BAWE Written All (8m)* sorted subcorp +assoc on left side for 10,000

2119 hits

Standardized to 311 per million (hits/corpus size x 1,000,000)

FREQUENCY BY SUB-CORPUS: sociology=540 law=290 english=189 anthropology=176 classics=172 history=114 linguistics=113 psychology=97 us_studies=78 politics=69 archeology=40 medicine=37 publishing=37 philosophy=34 mathematics=29 health=28 business=22 biology=15 hospitality=13 economics=9 food_science=8 agriculture=4 cybernetics=2 planning=2 engineering=1 architecture=0 chemistry=0 computer_sci=0 meteorology=0 physics=0

RANGE: family sex appears in 25 out of 30 sub-corpora (range=83.33%)

Click any [KEYWORD](#) for more context

001. ck cross sows in rotary field system 16,000 mixed [SEX](#) free range broiler chickens 180 South Devon an [agriculture]
002. led by the breed, the type of nourishment and the [SEX](#). Above is a drawing of a cow and the parts of [agriculture]
003. rters when fully matured. Manydown could use both [SEXES](#) for meat production, and also sell the large [agriculture]
004. Field A., Discovering Statistics Using SPSS (and [SEX](#), drugs and rock'n'roll), second Edition, SAGE [agriculture]
005. ions associated with contact. Thus we can see how [SEXUAL](#) division, class division, moral ideals and [anthropology]
006. sis for social definitions. What cultures make of [SEX](#) differences is almost infinitely variable, so [anthropology]
007. s to be humorous; homely, while a Geisha is to be [SEXY](#), etc. While this image does not appeal to the [anthropology]
008. at this and establish a clear distinction between [SEX](#) and gender. During this time it became clear t [anthropology]
009. reviously, the biological differences between the [SEXES](#) could not be used as a universal determinant [anthropology]
010. alternative theory that they propose is that both [SEX](#) and gender are socially constructed (Moore, 19 [anthropology]
011. e of socially constructed understandings of them. [SEXUAL](#) intercourse and human reproduction are not [anthropology]
012. opose that there is a further distinction between [SEX](#) and gender: 'Sex', represents a particular con [anthropology]
013. is a further distinction between sex and gender: '[SEX](#)', represents a particular construct of human b [anthropology]
014. ruct of human bodies, particularly Euro-America. [SEX](#), refers to the physical nature of human bodies [anthropology]
015. gender, relates to what different cultures see as [SEX](#) (Moore, 1994). One fact that must be indicated [anthropology]
016. indicated is that every culture cannot understand [SEX](#) without 'sex' (their own definition); this can [anthropology]
017. that every culture cannot understand sex without '[SEX](#)' (their own definition); this can therefore be [anthropology]
018. If even the biological definitions are variable, [SEX](#) cannot exist as a universal concept. Gender or [anthropology]
019. oncept. Gender or the socially applied meaning of [SEX](#) has the prevailing role and consequently gende [anthropology]
020. ale geladas have large shaggy manes of hair. Both [SEXES](#) have a naked bright pink hourglass shape on [anthropology]
021. ked bright pink hourglass shape on their chest, a [SEXUAL](#) displays such as estrus in females is signa [anthropology]
022. ividuals rear feeding have moved the necessity of [SEXUAL](#) display to the chest of the individual. The [anthropology]

2. Second-Sort direct from output

- E.g., Frequency of family members

2103. rly childhood are not sexually attracted to each other when SEXUALLY mature, this explains partly why humans rarely mate [anthropo]
2104. being too physical, too hard. This black masculinity is a 'SEXUALLY defined masculine ideal rooted in physical dominati [sociology]
2105. male could also imply that the female is either inferior or SEXUALLY promiscuous for example master/mistress, bachelor/s [linguist]
2106. d in a way that females may have more experience with being SEXUALLY harassed than males, and as a result, have schemas [psycholog]
2107. is a higher demand for prostitution and thus the spread of SEXUALLY transmitted diseases such as AIDS. Housing provisio [planning]
2108. ale publications played upon the popular images of women as SEXUALLY insatiable, nagging scolds, as well as playing upon [history]
2109. the contradictions within it; for example, that women were SEXUALLY uncontrollable, yet the moral guardians of society. [history]
2110. respectability, money, and class to declare sisterhood with SEXUALLY disgraced women' in an attempt to relieve tensions [english]
2111. e, thus women develop an instinct for nurturing and men are SEXUALLY promiscuous to provide the best chances for their g [sociology]
2112. to 885,000 by the next period. SUSI competes with SAFE and SEXY (segment market leader) in the Singles market, and SAMA [business]
2113. whereas Encolpio is a feminine male that opposes the ideal, SEXY gladiator. The spectacle of the male, therefore, is com [classics]
2114. ist ideal of the independent women however these women were SEXY. Connell in Segal 1997: 60 Richardson D in Holland and [sociology]
2115. a Geisha is to be humorous; homely, while a Geisha is to be SEXY, etc. While this image does not appeal to the average W [anthropo]
2116. % in the first half of 1999. The use of the "funny, useful, SEXY" formula is apparently one which works for the 18-30 ag [publishin]
2117. le that have survival value. This is also described as the 'SEXY sons' hypothesis, Eysenk (2001). The female looks for a [psycholog]
2118. evi, 1996, p79). The females in this anime are very strong, SEXY and dont seem to wear much clothing, this may be to hel [anthropo]
2119. agidae taxon. In a scientific competitive climate the mos' SEXY' projects are the ones most likely to get a piece of th [anthropo]

Distribution for family sex: sex=890 sexual=733 sexuality=305 sexes=83 sexually=55 sexist=17 sexism=10 sexy=8 sexing=5 sexualised=5 sexed=3 sexless=1 sexualisation=1 sexualise=1 sexualising=1 sexualized=1

3. Click key for larger context

Barbara Ehrenreich and Deirdre English argued that the witchcraft phenomenon 'was a ruling class campaign of terror directed against the female peasant population...' However, recent historiography has strayed from this simplistic line of argument and developed many other reasons for the high numbers of women in witchcraft prosecutions. As Brian Levack argues women were more likely to be accused of witchcraft but there was nothing that excluded men. This point is interesting, as although there were many more female than male witches, men were still prosecuted and found guilty, which does not fit into Ehrenreich's and English's thesis. **One of the most common arguments in recent years has been that '...witchcraft was SEX - related but not a SEX -specific crime'.** Understanding the witchcraft phenomenon as a 'persecution of women'

4. Separable collocation

Re-Conc family sort BAWE Written All (8m)* sorted left +assoc out on right

16 hits
Standardized to 2 per million (hits/corpus size x 1,000,000)

FREQUENCY : sort=15 sorts=1

Click any [KEYWORD](#) for more context

- 001. he EAP support class as being there to ' [SORT out](#)' all the language problems of t [linguistics]
- 002. her brother was also angry and told to " [SORT yourself out](#) you skinny bitch". Aft [medicine] ←
- 003. mments to self (made by student), e.g. " [SORT out](#) quote and footnote". 2 2004-10 [english]
- 004. [out](#), grow up, want, upset, stop, bury, [SORT out](#), fester, perceive, make, feel, [linguistics]
- 005. chemical shift. The PENDANT carbon NMR [SORT](#) the signals [out](#) into how many hydro [chemistry]
- 006. s and thus does suggest an influence of [SORTS](#). As Kühl points [out](#), eugenicists [history]
- 007. what constituted reality requires some [SORT](#) of [outside](#) existence to the body, w [psychology]
- 008. ater galago karyotypes in an attempt to [SORT out](#) ambiguities regarding interspec [anthropology]
- 009. can generally depend on the business to [SORT out](#) the problems quickly and to the [business]
- 010. e of its whole legal system if it is to [SORT out](#) this crises, and even that may [law]
- 011. ersity of Natal, South Africa, aimed to [SORT out](#) the Galagidae current phylogene [anthropology]
- 012. ites in Central Africa in an attempt to [SORT out](#) the taxonomic ambiguities of th [anthropology]
- 013. ntries across the world, which aimed to [SORT out](#) their own problems without Amer [politics]
- 014. s around the cement mixer. And I try to [SORT out](#) what has happened to me. The bu [english]
- 015. literature review of each species will [SORT out](#) what type of research that has [anthropology]
- 016. isible hands', in which the market will [SORT an unethical behaviour out](#), thus pr [agriculture] ←

5. Sort by VP level

Most comprehensible contexts first ...

- ast to Schopenhauer, who, "has made it [pity] the [VIRTUE](#)" (A7) and for whom no action could better d [1.00]
- ly from the outside; rather it sublates itself by [VIRTUE](#) of its own nature, and passes over, of its [1.00]
- e. As he writes, "this necessary perspectivism by [VIRTUE](#) of which every center of force - and not o [1.00]
- been born since Socrates' death in 399BC. Either [VIRTUE](#) is teachable or it is not, and it would be [1.00]
- roblem as, at first while it would only 'sap' the [VIRTUES](#) of public life, in the long run would dest [1.00]
- eone were to have a true knowledge of the form of [VIRTUE](#), he could teach it, but as no one can possi [1.06]
- but nobody knows how to do it. The lower form of [VIRTUE](#), on the other hand, has nothing to do with [1.06]
- question whether it is fair to connect these two [VIRTUES](#). In the end it would appear to me that her [1.06]
- that is ever in progress. Man changes himself by [VIRTUE](#) of his very nature. The form of man is his [1.06]
- lls us that Plato 'intended us to understand that [VIRTUE](#) can be taught, but that it is not taught at [1.07]
- but that it is not taught at present because such [VIRTUE](#) as exists at present is not knowledge.' (53 [1.07]
- ust believe that for a man to be entirely without [VIRTUE](#) is certainly possible, and that man would n [1.07]
- tain peace rather than war (another very Augustan [VIRTUE](#), a love of peace), and he cares for his men [1.08]

- l., 2004). References Anderson, K. (2005). On the [VIRTUES](#) of Multilateral Trade Negotiations. The Eco [2.43]
- available in certain tortuous strict liability by [VIRTUE](#) of legislation, this is usually confined to [2.45]
- eans things like habits or mannerisms - but where [VIRTUE](#) 'that vaguely-apprehended something that ma [2.45]
- and die for the French Republic is ennobling and [VIRTUOUS](#), and thus elevates the Revolution. In do [2.46]
- e Rome the imperialistic oppressor of implausibly [VIRTUOUS](#) and racially harmonious groups." The pre [2.50]
- lic , Oxford: Oxford University Press, p. 113 The [VIRTUE](#) of Temperance is somewhat ambiguous as it [2.56]
- to laugh at the other Sophists who claim to teach [VIRTUE](#). (Plato 95c) scorns those who claim to tea [2.62]
- ssions was amour the soi (love of self), of which [VIRTUE](#) and piti é were manifestations within soci [2.78]
- s view. Mina Harker embodies all the good womanly [VIRTUES](#), while Lucy Westenra is a voluptuous bloo [2.80]
- (Ithaca/London, 1986), p. 218. Blum, Republic of [VIRTUE](#) , p. 32. Hampsher-Monk, 'Rousseau and total [3.00]
- the zoon logon exon . The formal cause is that by [VIRTUE](#) of which the end is imprinted on the begin [3.21]
- ore independent and lacking traditional Victorian [VIRTUES](#). However, Mina Harker does in fact embody [3.38]
- d problems persist for those who classify taxa in [VIRTUE](#) of their monophyletic basis, in that; there [3.77]
- till expose Cecilia to public disapprobation. Her [VIRTUE](#) is redeemed somewhat because Pimienta, the [3.78]

6. With easy extraction for worksheet etc.

Concordance extract for *family* **VIRTUE** With no associate on EITHER side sorted level

- 002. Copenhagen, who, "has made it [pity] the [VIRTUE](#)" (A7) and for whom no action coul [1.00]
- 003. of all three in so far as they are the [VIRTUES](#) of a functionary unity."(Annas, [1.00]
- 005. ites, "this necessary perspectivism by [VIRTUE](#) of which every center of force - [1.00]
- 007. Philosophy 0215 Does the Meno show that [VIRTUE](#) is teachable? The question about [1.00]
- 008. inion (doxa), the more often we will be [VIRTUOUS](#). After all, as Guthrie says, 'd [1.00]
- 010. since Socrates' death in 399BC. Either [VIRTUE](#) is teachable or it is not, and it [1.00]
- 011. olve the problem of the teachability of [VIRTUE](#) is, but rather to demonstrate in [1.00]
- 013. the Brazilian blacks became a national [VIRTUE](#). In Mexico, the corrido and new p [1.00]
- 014. to have a true knowledge of the form of [VIRTUE](#), he could teach it, but as no one [1.07]
- 016. t is not taught at present because such [VIRTUE](#) as exists at present is not knowl [1.08]
- 017. whether it is fair to connect these two [VIRTUES](#). In the end it would appear to m [1.08]

User selected output: 11 selected from 394 available in Corpus=bawe_complete on lextutor.ca/conc/ set to 9999 max on 2019/5/28

7. Disambiguation with collocation

[Home](#) > [Concordancers](#) > Corpus search input (Eng)

Corpus Concordance English v.8

» With sub-corpus sort on * asterisked * corpora +NEW* BAWE Brit Academic Written (8 million/30 subs) » [DEM](#)

[French](#) [German](#) [Spanish](#) English

Base Speed \cong 1 second per million words of corpus

Add more for extras (associated words, family search, sub-corpus)

Keyword(s): (Max chars.) In corpus:

OPTION : With associated word(s) within words

and **NONE** of these words anywhere in the line

CONTROLS : **NEW SORT** ↓ Sort Lines by Mean VP ([Demo](#))

Sorted By of keyword | or | Line Width Number of Lines Gapped

On STARTS or FAMILY searches, "sort by keyword" gives distribution of forms before concordance lines (+ again at bottom)

+ Scan for any recurring word (potential collocate) within words presenting \leq times

(Where would such a list come from?)

- 1315. n in early 2003. Without [MONEY](#) for advertising, tr [politics]
- 1316. e reason to do any work: [MONEY](#). This view has been [business]
- 1317. 0's, the supply of world [MONEY](#) fluctuated in conve [economics]
- 1318. spiritually aware world, [MONEY](#) creates a calculati [sociology]
- 1319. 77)"Governing the Worlds [MONEY](#); Don't Just Do Some [economics]
- 1320. dcare) work is not worth [MONEY](#), is therefore value [sociology]
- 1321. lding payment. 'X owes Y [MONEY](#)' is a summary of ma [philosophy]
- 1322. that they will lend you [MONEY](#), your maxim therefo [philosophy]

Extract checked ↓ items: All | | [any10](#) | [20](#) | [30](#) | [50](#)

LEFT IMMEDIATE 1 WD COLLOCS (Freq>=2, with left/right x 1-2-3 wd sort) for **money** (assoc=):

of=264 the=154 and=68 for=65 more=38 =31 that=21 their=18 in=16 much=14 to=14 less=13 make=13 save=13
as=9 enough=9 this=9 anti=-7 nis=7 invested=7 public=7 real=7 which=7 any=6 borrow=6 saving=6 when=6
between=5 borrowing=5 by=5 domestic=5 have=5 little=5 lose=5 spend=5 them=5 time=5 bank=4 dirty=4 ext
than=4 without=4 ...=3 3:=3 against=3 anti=3 but=3 costing=3 earn=3 high-powered=3 increases=3 increa
paying=3 politics=3 raise=3 spending=3 (=2 -=2 2005=2 are=2 borrowed=2 boyfriends=2 cnn=2 creating=2
farm=2 fixed=2 get=2 greater=2 how=2 if=2 increase=2 international=2 into=2 invest=2 just=2 liquid=2

SUMMARY : KEYWORD=MONEY; NUMBER=1322; LANGUAGE=English; SEARCHTYPE= equals; SORT=1 words left of keyword; C
AssocSide=/SIDE= either ; GAPPED=no_gaps; LINEWIDTH=50 chars; DEMO=no; CPU SECS=3.44

19 hits

Standardized to 3 per million (hits/corpus size x 1,000,000)

FREQUENCY BY SUB-CORPUS: agriculture=7 archeology=2 architecture=2 english=2 food_science=
 planning=1 sociology=1 anthropology=0 biology=0 business=0 chemistry=0 classics=0 computer
 cybernetics=0 economics=0 engineering=0 health=0 history=0 hospitality=0 law=0 linguistics=
 mathematics=0 medicine=0 meteorology=0 philosophy=0 physics=0 psychology=0 publishing=0 us

RANGE: family *bank* appears in 8 out of 30 sub-corpora (range=26.67%)

Click any [KEYWORD](#) for more context

- 001. low direction of the stream on the left [BANK](#) from which the flow feeds the grazi [agriculture]
- 002. azing land. While the slope of the left [BANK](#) is relatively constantly flat, that [agriculture]
- 003. ntrolled with sandbags piled up at both [BANKS](#) as the river reaches the mill. BAN [agriculture]
- 004. th banks as the river reaches the mill. [BANK](#) TYPE: Vertical or sloping earth cli [agriculture]
- 005. was very steep. ADJACENT LAND USE: Left [BANK](#): The land was under rough grazing, [agriculture]
- 006. nd *S. cinerea* . RB: Similar to the left [BANK](#), with more diversity of flora, incl [agriculture]
- 007. ant during rearing of juveniles. Beetle [BANKS](#) that join the River Thames to the [agriculture]
- 008. tlement. Woodland areas, river and lake [BANKS](#), and wetland areas such as the Som [archeology]
- 009. of the tombs. They are also on the same [BANK](#) of the river, spaced a fair distanc [archeology]
- 010. ated on the east side of Oxford, on the [BANK](#) of the River Cherwell. The college [architecture]
- 011. n belt and within 50m meters of a river [BANK](#), 500m of Ancient Woodland, 25m of a [architecture]
- 012. the shimmering crane bird on the river [BANK](#) White like fresh milk! His roar is [english]
- 013. the shimmering crane bird on the river [BANK](#)', here the line breaks without inst [english]
- 014. Vosges#. In the Rhine valley, the west [BANK](#) faces the east and so is warmed by [food_science]
- 015. is warmed by the morning sun. The west [BANKS](#) receive the suns rays for the maxi [food_science]
- 016. wars such as in the Gaza Strip and West [BANK](#) where people were forcibly removed [planning]
- 017. e, Robert (2001) 'Showdown at the World [BANK](#)', New Left Review 7, Accessed on 08 [politics]
- 018. ion over Israeli occupation of the West [BANK](#), there are more significant problem [politics]
- 019. at Kind of Women: Stories from the Left [BANK](#) and Beyond . London: Virago Russell [sociology]

363 hits

Standardized to 53 per million (hits/corpus size x 1,000,000)

FREQUENCY BY SUB-CORPUS: politics=139 economics=68 agriculture=49 law=40 hospitality=34 business=7 publishing=4 history=3 us_studies=3 sociology=2 computer_sci=1 cybernetics=1 planning=1 anthropology=0 archeology=0 architecture=0 biology=0 chemistry=0 classics=0 food_science=0 health=0 mathematics=0 medicine=0 meteorology=0 philosophy=0 physics=0 p

RANGE: family bank appears in 15 out of 30 sub-corpora (range=50.00%)

Click any [KEYWORD](#) for more context

- 001. many internal donors, such as the World [BANK](#), which has just released a credit o [agriculture]
- 002. cept to Agricultural Innovation , World [BANK](#) Project Presentation in India Harlo [agriculture]
- 003. h Working Paper Series, No. 3332, World [BANK](#), [WWW] , Accessed in 25th November [agriculture]
- 004. ng Countries: How Much Is Enough? World [BANK](#) Policy Research Working Paper 3456 [agriculture]
- 005. g-run (Stern, 1999). Accordingly, World [BANK](#) advocated the discouragement of the [agriculture]
- 006. s from regional trade agreements. World [BANK](#) Policy Research Working Paper no. W [agriculture]
- 007. arch Working Paper no. WPS 1816 , World [BANK](#) Foroutan, F. (1998). Does membershi [agriculture]
- 008. untry more or less protectionist? World [BANK](#) Policy Research Working Paper no. W [agriculture]
- 009. arch Working Paper no. WPS 1898 , World [BANK](#) Gaisford, J.D., Kerr, W. A., Perdik [agriculture]
- 010. ent Research", Washington DC: The World [BANK](#). Gallaway, M.C., Blonigen, B.A., Fl [agriculture]
- 011. 002). A report carried out by the World [BANK](#) (2001) showed these afore mentioned [agriculture]
- 012. everity of the problem comes from World [BANK](#) (2001) where 23.4 billion tonnes of [agriculture]
- 013. vers and from cultivatable land, (World [BANK](#), 2004). This is not only a problem [agriculture]
- 014. Loess Plateau with help from the World [BANK](#) aimed to improve the potential for [agriculture]
- 015. talled sediment control dams. The World [BANK](#) played an important role in the pro [agriculture]
- 016. nd the initial investment lasts, (World [BANK](#), 2004). As was mentioned earlier, p [agriculture]
- 017. UN is involved together with the World [BANK](#) to help China in this period of tra [agriculture]
- 018. cal scale, with the help from the World [BANK](#) and FAO, enabling farmers to take c [agriculture]
- 019. den University Press, Netherlands World [BANK](#) (2001). China: air, land, and water [agriculture]
- 020. pment Programme was financed from World [BANK](#) funds. PHARE programmes were the fi [agriculture]
- 021. ors for educational projects, the World [BANK](#), has been currently funding 157 pro [agriculture]
- 022. nd Keynan (2000) produced for the World [BANK](#) shows that paid extension is "feasi [agriculture]
- 023. Technology Transfer in Nicaragua, World [BANK](#), [WWW] , Accessed in October 10th 2 [agriculture]
- 024. Accessed in October 14th 2005. World [BANK](#) (2005). Participation in Agriculture [agriculture]

7. Challenging gap worksheets

34 hits Hits/million cannot be calculated if Number of Lines user has chosen is set below the number a
FREQUENCY BY SUB-CORPUS: medicine=20 english=6 economics=1 engineering=1 health=1 his
politics=1 sociology=1 us_studies=1 agriculture=0 anthropology=0 archeology=0 archite
business=0 chemistry=0 classics=0 computer_sci=0 cybernetics=0 food_science=0 hospita
linguistics=mathematics=0 meteorology=0 physics=0 planning=0 psychology=0 publishin
RANGE: family xxx appears in 10 out of 30 sub-corpora (range=33.33%)

Click any KEYWORD for more context

- | | | | |
|------|--------------------------|--|---|
| 001. | <input type="checkbox"/> | e and the threat of hostile takeover is _____ | (b) large groups of firms are i [economics] |
| 002. | <input type="checkbox"/> | of 3 mm to a final diameter of 1 mm are _____ | greater than, for instance, dra [engineering] |
| 003. | <input type="checkbox"/> | otif in the final stanza ensures that a _____ | sense of loss and bereavement p [english] |
| 004. | <input type="checkbox"/> | tation." As with 'The Tell Tale Heart', _____ | instability and distorted reali [english] |
| 005. | <input type="checkbox"/> | nt of view; so shallow, so delusive, so _____ | .in every other." Images of d [english] |
| 006. | <input type="checkbox"/> | ustration thus making dissatisfaction a _____ | force; Marlow's tangible disill [english] |
| 007. | <input type="checkbox"/> | e narrator's presence, intensifying the _____ | , melancholic despondency that p [english] |
| 008. | <input type="checkbox"/> | eeing air was stunned' (1.9) engenders a _____ | aura of paranoia. Depicting a b [english] |
| 009. | <input type="checkbox"/> | f the questions to be asked there was a _____ | imbalance of power, with the in [health] |
| 010. | <input type="checkbox"/> | ty, the sublime landscape, unity with a _____ | nature, and Pantheism, all of w [history] |
| 011. | <input type="checkbox"/> | ed pain. Synovium. The synovium was not _____ | . Muscle bulk. The leg was swoll [medicine] |
| 012. | <input type="checkbox"/> | r. Distension/hepatomegaly/spleomegaly/ _____ | faeces in colon. BS +. Respirat [medicine] |
| 013. | <input type="checkbox"/> | ves on palpation. The apex beat was not _____ | . No pain was elicited on palpat [medicine] |
| 014. | <input type="checkbox"/> | No left or right parasternal heaves or _____ | thrills. The apex beat was loca [medicine] |
| 015. | <input type="checkbox"/> | patient was supine Palpation: Liver was _____ | five fingers breadth below the [medicine] |
| 016. | <input type="checkbox"/> | th below the subcostal marginSpleen was _____ | to the umbilicusNo other masses [medicine] |
| 017. | <input type="checkbox"/> | ilicusNo other masses felt. Kidneys not _____ | . Percussion: Liver was percusse [medicine] |
| 018. | <input type="checkbox"/> | scites, there was no loin tenderness or _____ | kidneys. Again, this diagnosis [medicine] |
| 019. | <input type="checkbox"/> | and lymphoedema. There is no history of _____ | cervical lymph nodes or yellowi [medicine] |
| 020. | <input type="checkbox"/> | qually on both sides. The apex beat was _____ | in the 5 th ICS MCL. No pain wa [medicine] |
| 021. | <input type="checkbox"/> | ess are also unlikely, as there were no _____ | masses or pain elicited on abdo [medicine] |
| 022. | <input type="checkbox"/> | nderness in the right iliac fossa and a _____ | lump may indicate the presence [medicine] |

RESULTS

Summary of feature use – all corpora

From **49,461** independent consultations since April 2018 , here are the proportion seeking...

- A particular corpus
- Search options (equals, family, starts, ends)
- Sort options (VP-Level, discipline, right, left)
 - Esp. Partitioned corpora
- Collocation
 - Fixed and separable
 - Esp. Disambiguation by multiple
- Larger context
- Output shaping
- Gapped output







Data restriction

The following data does **not** include these :

- Concordance DEMO routines from input page
 - $\cong 20$ in number and probably $>50\%$ of searches
- Automated concordance calls from other routines
 - With pre-set defaults

C
O
R
P
U
S

C
H
O
I
C
E

CORPUS	SIZE (wds)	hits	% of 542,777
Brown	1 m	189,272	34.87
BNC_legal	2.3 m	58,723	10.82
<u>Brown_BNCW*</u> 	2 m	41,339	7.62
<u>BNC_sp_v_wrt*</u> 	2 m	41,339	7.62
Brown_BNCW	2 m	22,904	4.22
Acada_corp	6 m	21,161	3.90
BAWE* 	8 m	19,749	3.64
bncoca_1-2	14 m	19,611	3.61
BNC Speech Sampler	1 m	19,162	3.53
BNC Medicine	1.4 m	17,191	3.17
BNC Spoken	10 m	15,500	2.86
BNC Written Sampler	1 m	15,171	2.80
BNC_humanities	3.4 m	12,790	2.36
TV Talk	.5 m	11,920	2.20
Wiki* 	1.1 m	10,639	1.96
BNC Commerce	2.2 m	8,740	1.61
Graded 2k	1 m	8,406	1.55
Graded 1k	1m	7,470	1.38
Dr House* 	.8 m	3,801	0.70
Graded* 	1 m	3,511	0.65

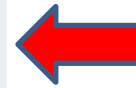
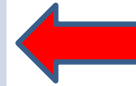


* = partitioned corpus, otherwise 1-file corpus

**S
E
A
R
C
H

T
Y
P
E**

SEARCH TYPE	# HITS	% of 542,777
EQUALS	478,611	88.18
FAMILY	48,891	9.01
STARTS	7,374	1.36
CONTAINS	5,927	1.09
ENDS	1,841	0.34



W
D
S.
S
E
A
R
C
H
E
D

) equals ▾ look up
With associated word(s)

family ▾ look
With associated word(s)

MULTIWORD SEARCH	# HITS	% of 542,777
1 WORD	373,062	68.7
2 WDS	57,879	10.66
3 WDS	67,751	12.48
4 WDS	6,082	1.12
5 WDS	2,318	0.43



SEARCH_WORD \w*.\w \w*.\w \w*.\w \w*.\w \w*.\w

**S
O
R
T

T
Y
P
E**

SORT TYPE	# HITS	% of 542,777
KEY (default)	257,153	47.4
LEFT	207,708	38.3
RIGHT	54,240	9.9
VP-LEVEL	3,864	0.7
SUB-CORPUS	1,248	0.2



[Home](#) > [Concordancers](#) > [English Input](#) [[«Back](#)] > [Eng Output](#) (Back keeps original settings) Copiable extract-Link to this data >

Concordance for *equals* sure in **bawe_complete** sorted 1 wd right of key

 Dictionary Eng_Eng Speak Eng-US

Extract: All | 0 | [any10](#) | [20](#) | [30](#) | [50](#)

Re-Conc
equals
sure
BAWE Written All (8m)*
sorted
1 wd
right
+assoc
make
on
left
side

O
U
T
P
U
T

S
H
A
P
I
N
G

REFIRE	% of 542,777
266,796	49.15

(Max chars.) In corpus:

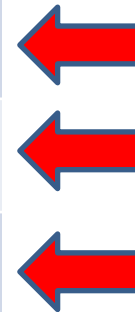
With associated word(s) within words to

and NONE of these words anywhere in the line

N
 U
 M.

 C
 O
 L
 L
 O
 C
 A
 T
 E
 S

# OF ASSOC_WORD	# HITS	% of 542,777
0 WORDS	465813	85.8
1 WORD	75314	13.9
2 WDS	4304	0.79
3 WDS	1883	0.35
4 WDS	749	0.14
5 WDS	460	0.08



ASSOC_WORD \w*.\w \w*.\w \w*.\w \w*.\w \w*.\w

Requests for larger contexts on two random days

Popular Pages Compleat Lexical Tutor - Thu 23rd May, 2019	
Hits	Title
974	VP 1-WORD (HOME)
907	Compleat Lexical Tutor
594	CORPUS CONCORDANCE OUTPUT <<<<<<
588	VP 1-WORD (EDIT-TO-PROFILE)
563	VOCABPROFILE COMPLEAT - INPUT
552	VOCABPROFILE COMPLEAT - OUTPUT
494	VOCABPROFILE ENGLISH
446	VOCABPROFILE ENGLISH OUTPUT
253	CORPUS CONC ENG
234	VP HOME
219	CORPUS CONCORDANCE OUTPUT <<<<<<
187	VST (computer/test)
178	CORPUS CONC ENG
170	TESTS HOME
153	ONLINE CONCORDANCERS
138	VST (computer/practice)
121	Compleat Lexical Tutor
116	VOCABPROFILE COMPLEAT - INPUT
115	VOCABPROFILE COMPLEAT - OUTPUT
98	FREQUENCY HOME
81	LEVELS TEST (PRODUCTIVE)
78	LARGER CONTEXT <<<<<<
73	HYPertext 2
69	TEXT-LEX COMPARE
64	FREQUENCY INDEXER ENGLISH
41	LARGER CONTEXT <<<<<<
40	404 Not Found

**119/813
= 15%**

Popular Pages Compleat Lexical Tutor - Tue 28th May, 2019	
Hits	Title
875	VOCABPROFILE COMPLEAT - INPUT
782	VP 1-WORD (EDIT-TO-PROFILE)
777	VOCABPROFILE COMPLEAT - OUTPUT
759	VP 1-WORD (HOME)
742	Compleat Lexical Tutor
585	CORPUS CONCORDANCE OUTPUT <<<<<<
464	CORPUS CONCORDANCE OUTPUT <<<<<<
438	VOCABPROFILE ENGLISH
427	VST (mobile/test)
325	CORPUS CONC ENG
314	VOCABPROFILE ENGLISH OUTPUT
286	CORPUS CONC ENG
266	VP HOME
210	TESTS HOME
163	VST (computer/test)
145	Compleat Lexical Tutor
140	ONLINE CONCORDANCERS
116	LEVELS TYPE TESTS
79	LEVELS TEST (PRODUCTIVE)
70	LIST_LEARN ENGLISH
66	VST (computer/test)
63	COVERAGE CALCULATOR
59	VOCABPROFILE COMPLEAT - OUTPUT
41	Vocab Stats
41	LARGER CONTEXT <<<<<<
40	LARGER CONTEXT <<<<<<
32	N-GRAM INPUT

**82/1049
= 8%**

Interesting gap-tasks?

- Negligible across corpora
- despite their popularity with learners

in my experience

>> and proven
learning power
>> particularly
in transfer
dimension

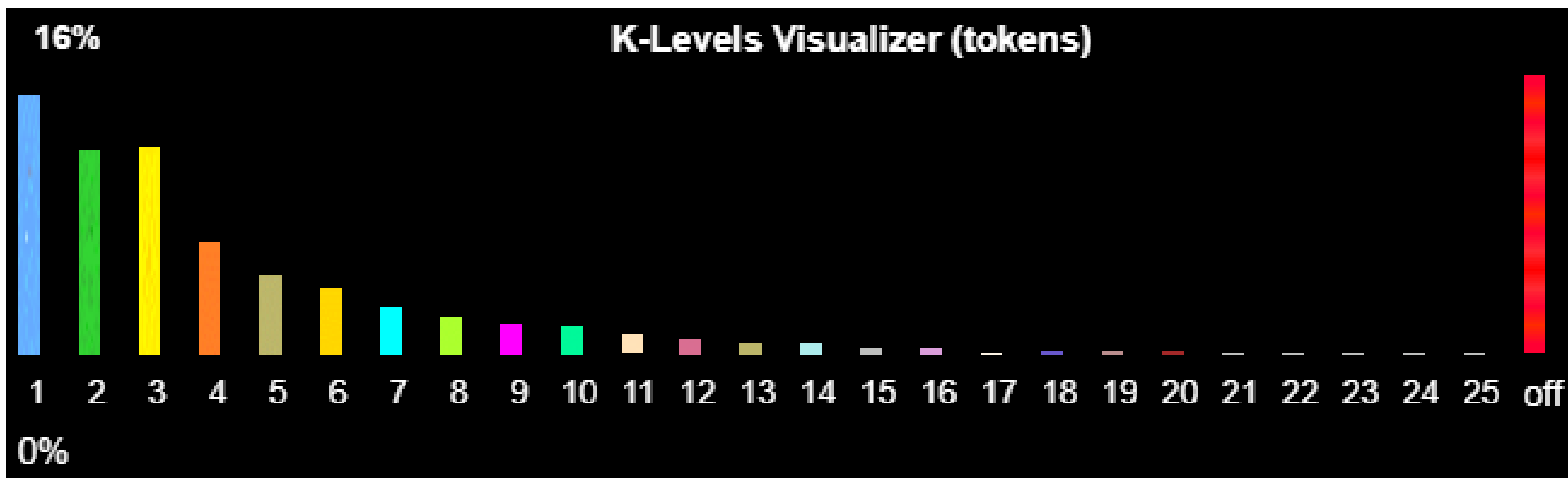
```
34 hits Hits/million cannot be calculated if Number of Lines user has chosen is set below the number a
FREQUENCY BY SUB-CORPUS: medicine=20 english=6 economics=1 engineering=1 health=1 his
politics=1 sociology=1 us_studies=1 agriculture=0 anthropology=0 archeology=0 archite
business=0 chemistry=0 classics=0 computer_sci=0 cybernetics=0 food_science=0 hospita
linguistics=0 mathematics=0 meteorology=0 physics=0 planning=0 psychology=0 publishin
RANGE: family xxx appears in 10 out of 30 sub-corpora (range=33.33%)
```

Click any **KEYWORD** for more context

```
001.  e and the threat of hostile takeover is _____ (b) large groups of firms are i [economics]
002.  of 3 mm to a final diameter of 1 mm are _____ greater than, for instance, dra [engineering]
003.  otif in the final stanza ensures that a _____ sense of loss and bereavement p [english]
004.  tation." As with 'The Tell Tale Heart', _____ instability and distorted reali [english]
005.  nt of view; so shallow, so delusive, so _____ .in every other." Images of d [english]
006.  ustration thus making dissatisfaction a _____ force; Marlow's tangible disill [english]
007.  e narrator's presence, intensifying the _____, melancholic despondency that p [english]
008.  eeing air was stunned'(1.9) engenders a _____ aura of paranoia. Depicting a b [english]
009.  f the questions to be asked there was a _____ imbalance of power, with the in [health]
010.  ty, the sublime landscape, unity with a _____ nature, and Pantheism, all of w [history]
011.  ed pain. Synovium. The synovium was not _____ . Muscle bulk. The leg was swell [medicine]
012.  r. Distension/hepatomegaly/spleomegaly/ _____ faeces in colon. BS +. Respirat [medicine]
013.  ves on palpation. The apex beat was not _____ . No pain was elicited on palpat [medicine]
014.  No left or right parasternal heaves or _____ thrills. The apex beat was loca [medicine]
015.  patient was supine Palpation: Liver was _____ five fingers breadth below the [medicine]
016.  th below the subcostal marginSpleen was _____ to the umbilicusNo other masses [medicine]
017.  ilicusNo other masses felt. Kidneys not _____ . Percussion: Liver was percusse [medicine]
018.  scites, there was no loin tenderness or _____ kidneys. Again, this diagnosis [medicine]
019.  and lymphoedema. There is no history of _____ cervical lymph nodes or yellowi [medicine]
020.  qually on both sides. The apex beat was _____ in the 5 th ICS MCL. No pain wa [medicine]
021.  ess are also unlikely, as there were no _____ masses or pain elicited on abdo [medicine]
022.  nderness in the right iliac fossa and a _____ lump may indicate the presence [medicine]
023.  as also been damaged. No tenderness was _____ over the fifth metatarsal, prox [medicine]
024.  s or deformities. HS I + II + 0 JVP → _____ periphera pulses bilaterally N [medicine]
```

Frequency of look-up words by VP

By k1=0-999 etc. by combined family count



Roughly

- 50% 1k-3k
- 25% 4k-25k
- 25% “off-list”

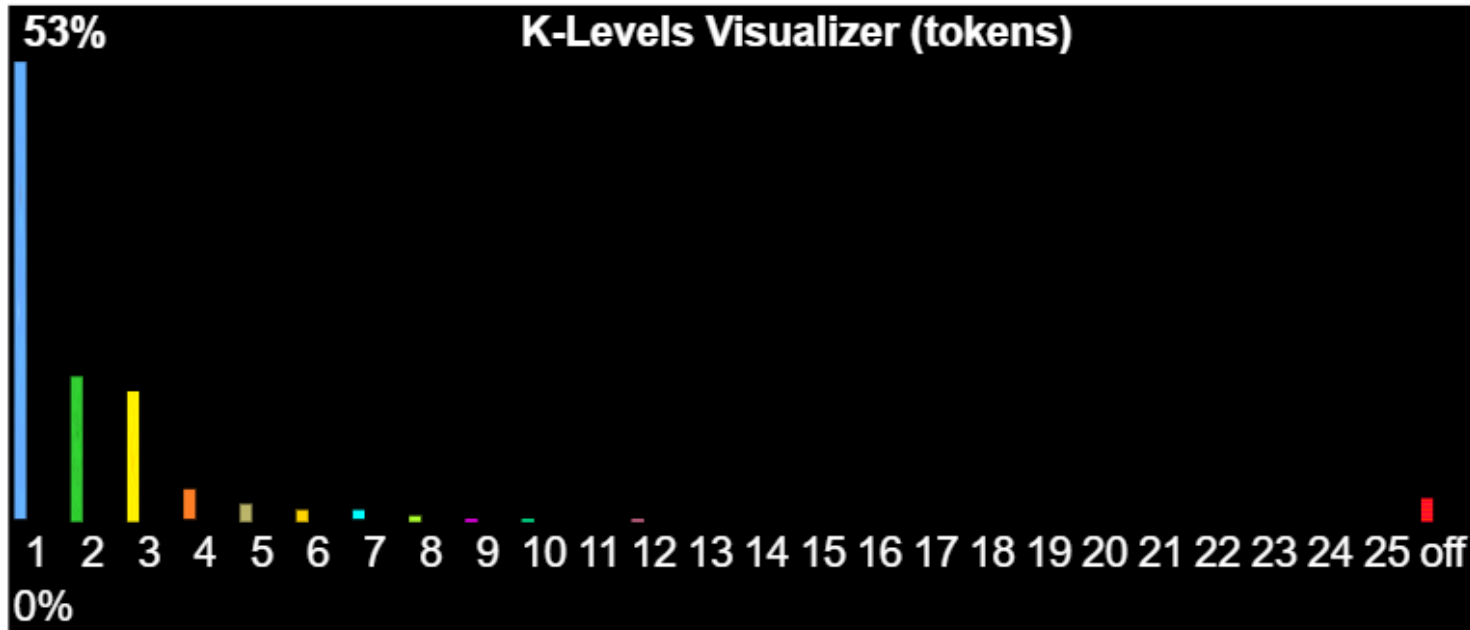
(mainly misspellings not found in VP’s BNC-Coca family lists of $25k \times 5 = 125,000$ word-forms)

And specifically BAWE Complete

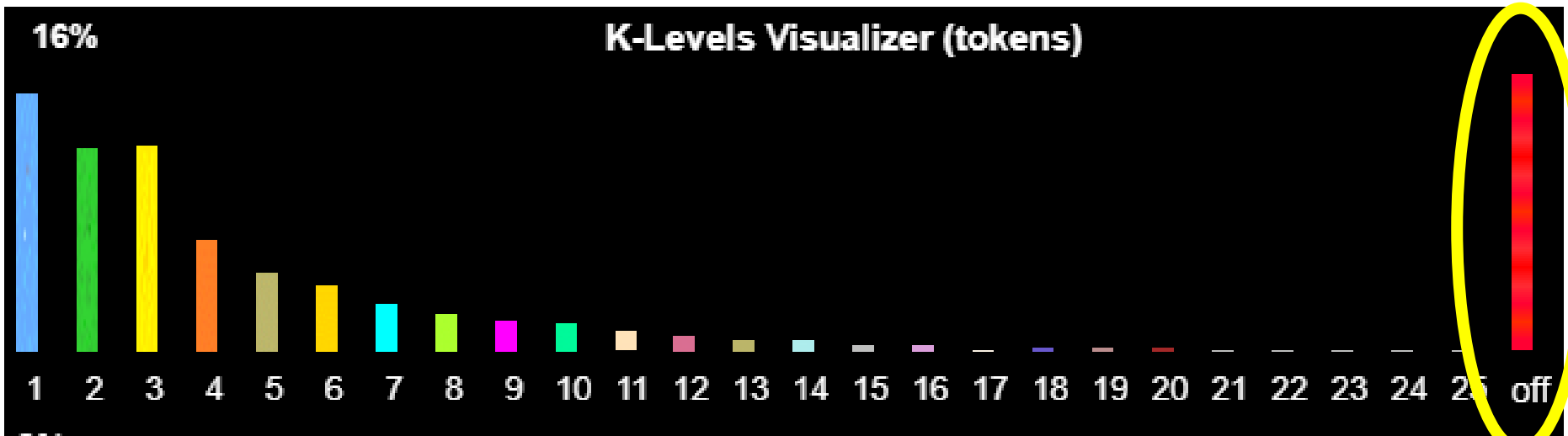
19,749 searches

	0	<u>%</u>	1	<u>%</u>	2	<u>%</u>	3	<u>%</u>	4	<u>%</u>	5	<u>%</u>
# SEARCH WORDS	-	-	12,763	<u>64.6</u>	3,307	<u>16.7</u>	1340	<u>6.8</u>	427	<u>2.2</u>	65	<u>0.3</u>
# ASSOC WORDS	16,453	<u>83.3</u>	2,834	<u>14.4</u>	9	0.0	0	0				
SEARCH TYPE	equals	%	family	%	starts	%	contains	%	ends	%		
	18,065	<u>91.5</u>	1,037	<u>5.3</u>	245	<u>1.2</u>	320	<u>1.6</u>	78	<u>0.39</u>		
SORT TYPE	key	%	left	%	right	%	VP-Level	%	subcorp	%		
	13,527	68.5	2,998	15.2	2,596	13.1	159	0.8	465	2.35		
SEARCH WORD	k1	%	k2	%	k3	%	k4	%	k5-25	%	OFF	%
K-LEVEL	15,404	<u>53.0</u>	4,900	<u>16.9</u>	4,417	<u>15.2</u>	1,049	<u>3.6</u>	2,181	<u>7.6</u>	1,085	<u>3.74</u>

Look-ups by frequency for BAWE Complete



For all corpora



What does frequency level of look-up items tell us?

- Hypothesis:
 - If look-ups are mainly 1k-3k, then the interest is usage characteristics of common words
 - Collocation, part of speech, register, how-to-use
 - If mainly 4k+, then meaning of uncommon words is the interest
 - Polysemy, homophony, inferencing from context, meaning within a discipline
 - **BAWE pattern suggests targeting of usage characteristics**

What does fewer off-list items tell us?

- Hypothesis:
 - Off-list items are mainly misspellings
 - High rate of misspellings suggests learners are working independently
 - Low rate suggests teacher has supplied a BAWE task
 - Which in turn supports the usage hypothesis

What distinguishes BAWE usage?

- Fewer off-list look-ups
- More common word look-ups
- → Hence focus on usage in teacher-led tasks

Everything else is the same as other corpora

- Low interest in corpus partitioning
 - search, sort, collocation options
 - broader context
 - collocation

Putting it all together

- Most users get one-word example from small, unpartitioned corpus
 - Minus interesting sort or collocation
- Interesting corpora + searches get low use
- Just enough to indicate...
 - They *are* discoverable
 - But that few discover them
- High quality corpus like BAWE makes minor difference

But who is this user ?



- If **weekend** is heavy, it is mainly students doing assignments
- If **weekday** is heavy, it is mainly teachers preparing and delivering in-class demos
 - Confirmed by emails

Hypothesis: Users are > 50% teachers

- The remainder are probably learners told to look at concordance by their teachers
 - But learners seem not adequately prepared to exploit resources most usefully

How many TESL training programs include *anything* about corpus ?

- How to search, interpret, generalize patterns, apply patterns...
 - Let alone get learner into DDL spin-offs as offered by Lextutor
- **Mini-Experiment:** Google the keywords “tesl training programs corpus DDL”
 - With the small problem that ‘DDL’ picks up ‘didactique des langues’

All

News

Images

Shopping

Maps

More

Settings

Tools

About 56'700 results (0.50 seconds)

TEFL/TESL Certificate Programs - Distance Learning Programs

<https://www.tesol.org/...your.../certificate-programs-and-distance-learning-programs> ▼

Those who decide not to begin their teaching career with a **TESL/TEFL/TESOL** degree or endorsement may choose to pursue a shorter term **certificate program**, ...

Missing: ~~corpus~~ | Must include: corpus

People also ask

How can I get my TEFL certification online for free? ▼

How do I get ESL certified? ▼

What qualifications do you need to teach English as a second language? ▼

What is a TESL certificate? ▼

Feedback

Accredited TESL Training Programs | TESL Ontario

www.teslontario.org/accreditation/tesl-training-programs ▼

The **TESL** Ontario accredited English teachers work for a variety of employers ... For information about the **TESL program** admission requirements, registration ...

Missing: ~~corpus~~ | Must include: corpus

T
O
P

2

S
E
A
R
C
H

R
E
S
U
L
T
S

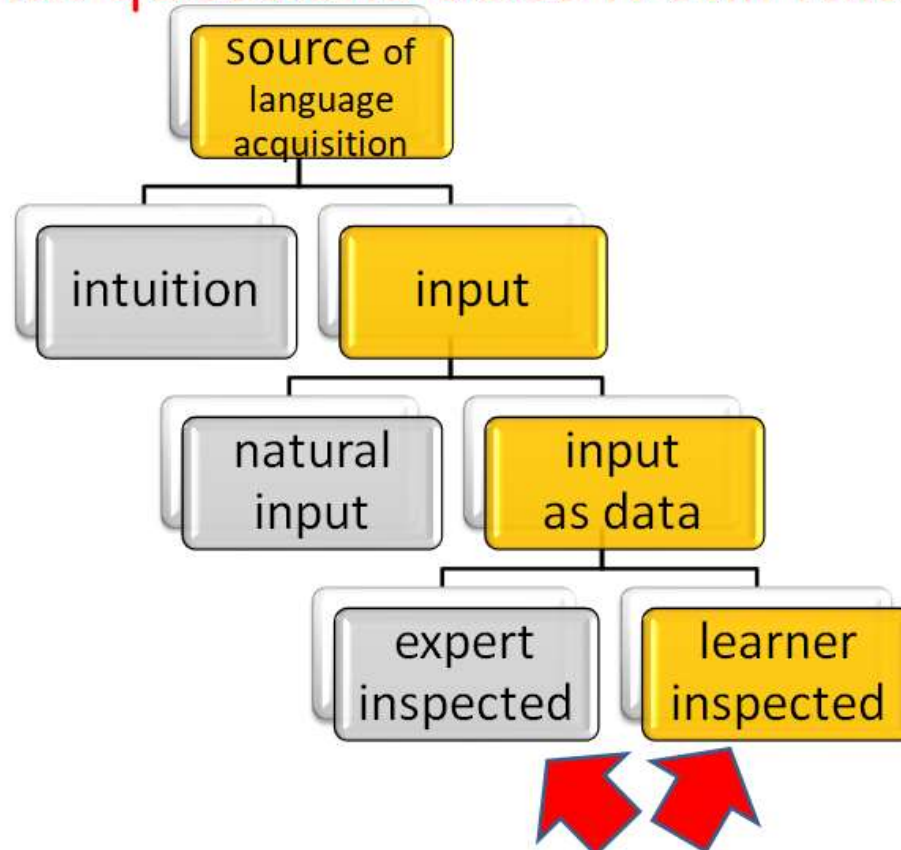
Conclusion

- Better corpus and search tools are not enough
 - Users will only scratch the surface
 - Whether learners or teachers
- Training in search & interpretation is essential
 - TESL MA programs are the logical place to start
 - But these have shrinking course loads, not expanding!
- The “corpus revolution” is reaching the classroom
 - But it is a minimal version

Conclusion

- So to answer the starting question

What questions does it still raise?



Conclusion

- DDL only makes sense if learners' feel they are making their own discoveries in the language data
 - In some sort of problem-solving framework
- But this takes some knowledge, preparation & planning to set up
 - I suspect we need the hidden hand of the expert a bit closer to the surface of DDL than it is at present
 - Make sure interesting data gets 'discovered'

Conclusion

- This hidden hand is needed to make sure that high quality corpora and search tools get used and are stretched
 - And people like us are pushed to make them even better
- But the Lextutor experiment shows this is unlikely to happen by itself

~END~