Data Driven Learning and the Arab learner : A particularly good match

Tom Cobb

Didactique des langues

Université du Québec à Montréal

cobb.tom.3@gmail.com

Saturday 5 Nov 2022

Abstract

Data-Driven Learning and the Arab Learner – A Particularly Good Match

ESL and ESP In the Arabian Gulf context have traditionally followed a top-down model, typified by dictionary definitions, grammar rules, and teacher-centered classrooms. My work in this setting, however, has shown that a different model is more likely to be successful: a bottom-up model typified by exemplar-based learning, usage-based pattern induction, and task-based problem solving. Since my early efforts with this approach in the 1990s at King Saud and then Sultan Qaboos universities, the approach has gone on to take shape internationally under the heading 'data driven learning,' and has been validated empirically in a wide-ranging meta-analysis (Boulton & Cobb, 2017). I will review the high points in this progression and argue that data-driven learning is of particular relevance to an Arabic speaking context because of (1) some features in the way written Arabic is processed, which are effectively training for data-driven learning; (2) the likelihood that in many Gulf settings, general ESL is likely to be a prelude to the study of ESP (English within a specific academic or professional domain), for which data driven learning is particularly well suited; and (3) the motivation within this learnership of a systematic basis in information technology.

What is Data Driven Learning (DDL)?

Language Learning involving the consultation of a *corpus*

Whether ~

1. Of a language as a whole
   - Ex Brown Corpus (1960s)
     - 1 million words
     - 500 textes on 15 topics of 2000 words
   - Ex BNC (1980s)
     - 100 million words
     - >4000 texts on 100 topics of 25,000 words
   - Ex COCA, Subtlex, TenTen …

2. Or the language of a course/curriculum
   - Texts + Media + Tests + Talk

Corpus – read and interpreted how ?

001. ☐	0 décembre, 134 personnes 53 hommes, 55 FEMMES 26 enfants de un à treize ans ét
002. ☐	de ces deux affaires, une homme et une FEMME : Alfred Sirven, ancien directeur
003. ☐	ns cent quatre-vingt-dix hommes et deux FEMM
004. ☐	chorégraphie. La pièce met en scène une FEMM
005. ☐	pide peur de l'homme, des hommes et des FEMM
006. ☐	août 1942. Déjà, de nombreux hommes et FEMM
007. ☐	amenant vers l'emploi des hommes et des FEMM
008. ☐	'hématocrite chez 46 athlètes hommes et FEMM
009. ☐	. Les hommes d'aujourd'hui traitent les FEMM
010. ☐	de vingt-cinq ans et aux hommes et aux FEMM
011. ☐	long desquelles se succèdent hommes et FEMM
012. ☐	ne se reproduisent, tous les hommes et FEMM
013. ☐	hommes d'affaires tout-puissants... "La FEMM
014. ☐	ravaillait comme un homme, ou comme une FEMM
015. ☐	onnage principal, l'homme trompé par sa FEMM
016. ☐	ports ambigus d'un vieil homme et d'une FEMM
017. ☐	janvier, traite ainsi des hommes et des FEMM
018. ☐	nt ce partage de la vie entre hommes et FEMM
019. ☐	emportent maintenant des hommes et des FEMM
020. ☐	e malade et doit subir la cruauté de sa FEMME Enlevé par ses hommes, il finit p
021. ☐	rieusement atteinte ? "J'espère que les FEMMES et aussi les hommes de ce pays e
022. ☐	vant ces ouvrages historiques signés de FEMMES et d'hommes publics, dont on imag
023. ☐	du Soudan ROUX OLIVIER DES hommes, des FEMMES et des enfants meurent de faim au
024. ☐	le monde a toujours fermé les yeux. Des FEMMES et des hommes n'ont cessé de tire
025. ☐	es", "filières criminelles") contre des FEMMES et des hommes prêts à se laisser

'Teachables' in a corpus output

Typical collocations of words
    ('hommes' before 'femmes')
Forms of words typically employed
    ('femmes' way more
    frequent than 'femme')

+ much more

4

The learning principle of DDL

Learners generalize from language data
- Rather than learn rules and apply them to data
- Using Bottom-Up not Top-Down processing

.. aided by computation
- Data assembled by computer software
- Patterns exposed by computer software
  - To different degrees for different linguistic phenomena
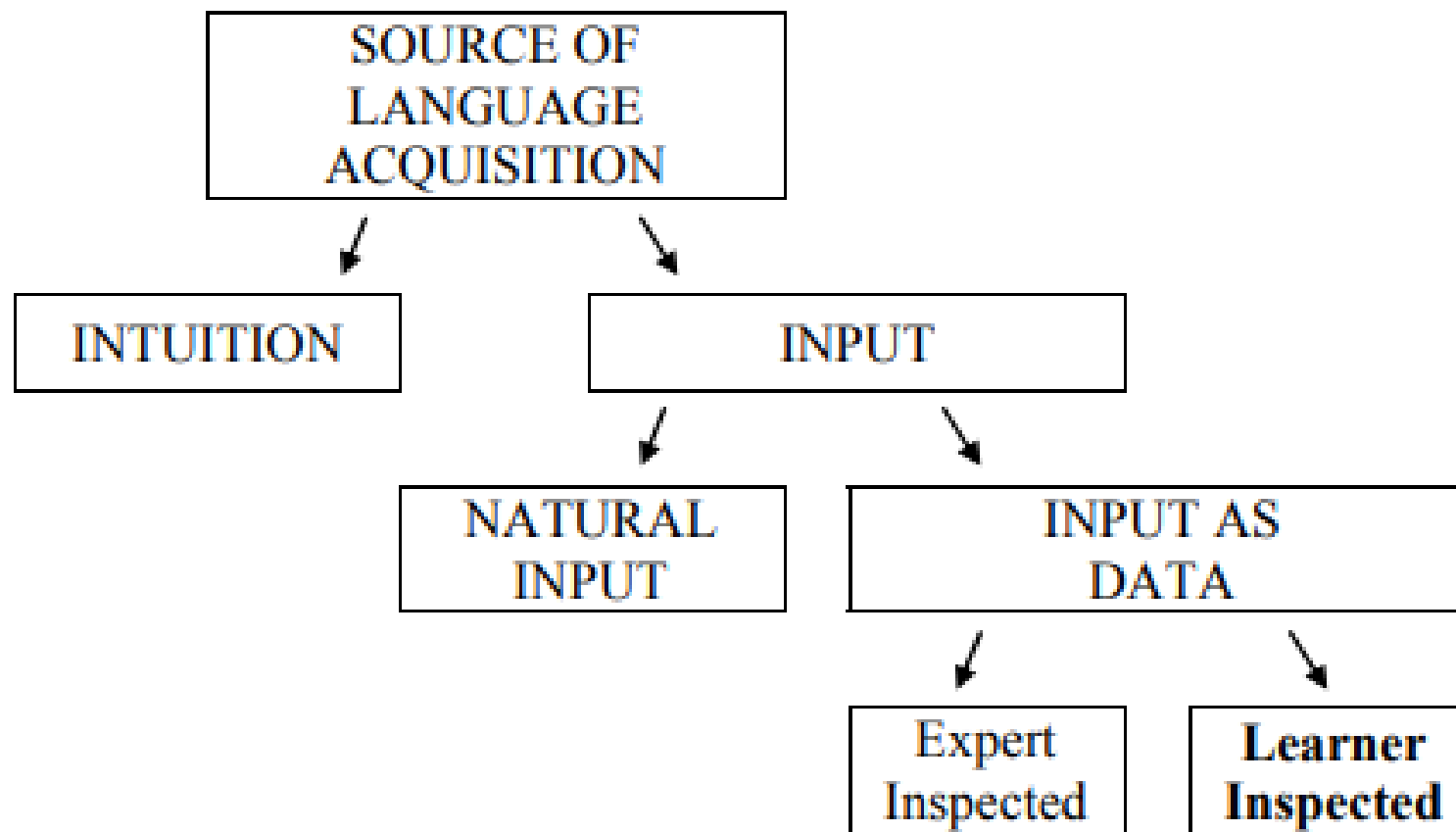
The place of DDL within SLA

Figure 1. The place of data-driven language learning in the broader scheme

Example of a DDL learning activity

Constructing word meanings from raw-ish **vs** from pre-processed data

←vs.→

Example of software making sense of language data

- Obvious example is TTS (text-to-speech)

- Algorithms pull the [ˈfaɪt] and [ˈnaɪt] out of *fight* and *night*
  - Makes the link between speech and text
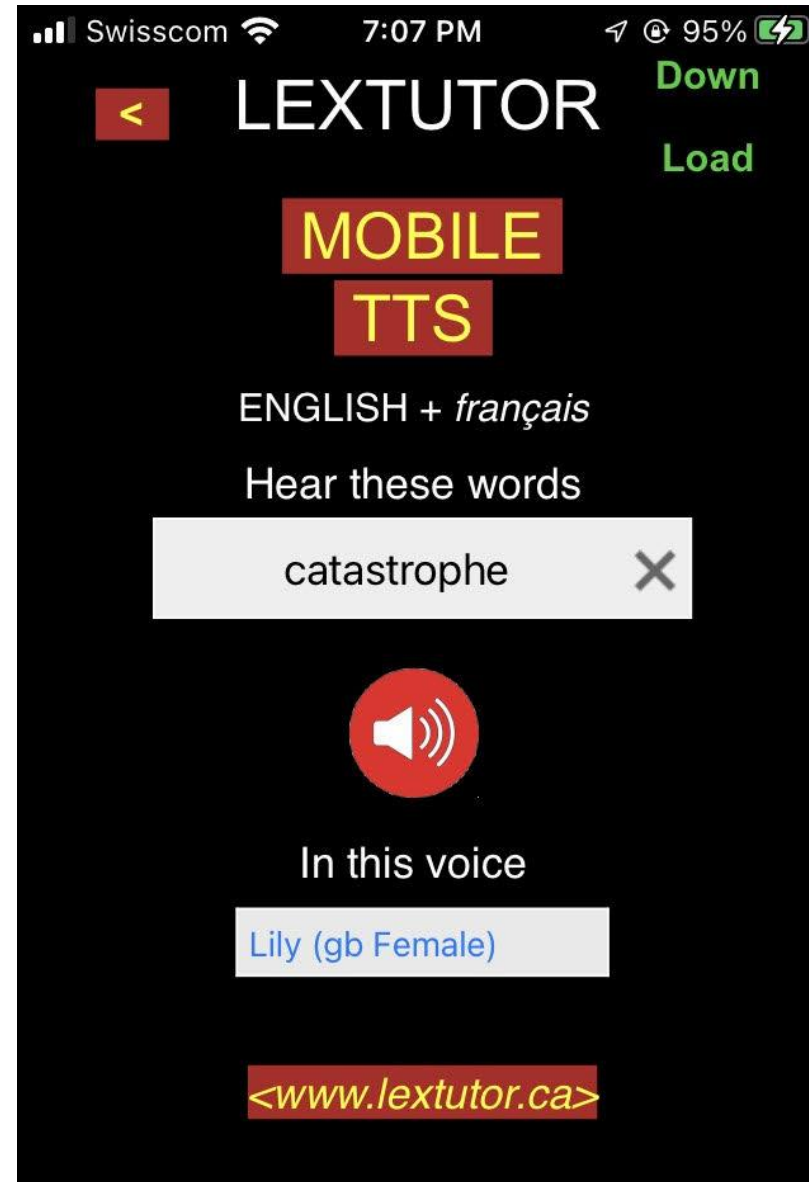
Example of software making sense of language data (2)

- Concordance lines sorted by VP (= frequency; more comprehensible first)

Home > Concordancers > English Input > Output   (« Back keeps original settings)   Copiable extract-link to this data ›
Concordance for *family* symptom in **house** ▉   Dictionary/TTS   Eng_Eng ▾ ▉

Extract checked ❗items: ☑All | Ø | any10 | 20 | 30 | 50   Go >

MODIFY:
family ▾ ⁝ symptom ⁝ Dr House* (811k) ⁝ **SORT** ▾ VP ▾ **ASSOC** ⁝ ON L|R ▾ FOR 25,000 ▾ L

017. ☐ It's a different situation, a different SYMPTOM, a different diagnosis. Differen [1.00]
018. ☐ him needing glasses. Apparently, it's a SYMPTOM. It means the condition predats [1.00]
019. ☐
020. ☐
021. ☐
022. ☐
023. ☐
024. ☐
025. ☐
026. ☐

710. ☐ e cirrhosis. Liver cysts have identical SYMPTOMS, with one addition, fever. And [4.12]
711. ☐ ia, depression, tachycardia. Withdrawal SYMPTOMS. Not applicable. The only side [4.17]
712. ☐ egative. Several meds could cause those SYMPTOMS. Tricyclic antidepressants, rit [4.20]
713. ☐ the delirium. But not the original lung SYMPTOMS. Legionnaire's could and cause [4.25]
714. ☐ aud's. Lupus is consistent with Emily's SYMPTOMS. Her lymph nodes aren't enlarge [4.25]
715. ☐ really change anything. Hemoptysis is a SYMPTOM of Behcet's. Except our treatmen [4.29]
716. ☐ . And less significantly, we have a new SYMPTOM. Anion gap acidosis. Who's chubb [4.30]
717. ☐ atient's religious extremism could be a SYMPTOM? A neuro disorder could explain [4.43]
718. ☐ pain, neurogenic bladder, psychosis. A SYMPTOM which we learned about when he t [4.44]
719. ☐ an nature. Gonorrhea explains all their SYMPTOMS. Fitzhugh-Curtis syndrome for h [4.50]
720. ☐ e. Hypervigilance, sudden irritability. SYMPTOMATIC of lunch with Cuddy? The pat [4.67]

9

VP = 1

Last few months, that hasn't been so easy. A guy got me the pills, and they control the stress. Look, don't say anything to Eddie, all right? Last thing I need is the guys thinking that I've gone soft. When was your last pill before we originally admitted you? Um The night before. Then we stuck you in here and you went into beta-blocker withdrawal. That's what caused the hypertension. When he got out of here, he took another pill, and the hypertension went away, which means it was never a SYMPTOM. We take that off the table, we've got vertigo, fever And stress. He's been in this game for a while, either he grew a conscience over the last few months, or he's producing excess adrenaline. Pheochromocytoma. What? That's Greek for "you'll be back doing bad things to good people in no time." MRI his adrenal glands. You're okay with that, right? You're not gonna put a pillow over his face? Oh, great, the

---

VP 5

Second half of my plan might need work. V-fib. Two rounds at 360 without establishing a rhythm. Wait! Dr Foreman! Clear. He's back. I sure hope this Alzheimer's drug trial is worth a lot to this hospital, 'cause we're wasting a crapload of time and man power just to let this guy watch his mind shrivel and keep his wife miserable. Heart attack while he was on interferon rules out encephalitis, but I still think I'm right about the neuro SYMPTOMS. It's probably aseptic meningitis. We would have seen meningeal signs. Given his abnormal mental status, not necessarily. Dr Foreman! Would you please reach under your chair'? It's just wet gum under here. Where's the card? What card? I'm happy to leave if you have a better theory than aseptic meningitis. Toxins. If this guy went on a walkabout, chances are he's done it before. Wife keeps the doors in the

Why *should* this learning approach 'work' ?

The case in principle (1)

Data-driven learning (DDL): using the [...] of CL
for teaching/learning/using a foreign/[...]

- Greater effort
  better retention/deeper learning

- Strategy training
  learn to observe patterns in environment (e[...]

- Incidental learning
  "pick up" collocations while assembling wor[...]

- Appeal to research-oriented clientele
  as opposed to ESL textbooks; PhD students,

The case in principle (2)

+4 related to Arabic speaking learner
- Lends itself to group problem solving tasks
- Does not encourage a surface memory approach
  - Rather encourages depth & transfer
- Can work at the processing level & target L1→L2 processing differences
- Allows practice of test formats (can re-randomize corpus data for tests)

Should work but *does* it ?

A 2017 meta-analysis compares DDL to a range of other ways of learning

Vocabulary
Multi-Word Units
Syntax
Grammar
Culture

LANGUAGE LEARN
A Journal of Research in Language
Languag

EMPIRICAL STUDY 😊 😊

Corpus Use in Language Lea
A Meta-Analysis

Alex Boulton and Tom Cobb
Université de Lorraine and Université du Québec a Montreal

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

With DDL 'winning' by an av. **effect size of 1.5** (by 1.5 std. deviations)

Ex. Control Group M= **65**, SD=15
     DDL Group   M= **87**, SD=15

12

None of the studies in the meta-analysis took place in the Gulf area

But they could have, because DDL and the Arabic speaking learner are a good fit

I will make my case with reference to a learning approach that was/is *not* a good fit

Namely the approach used in many Gulf language centers in their pre-academic ESL/ESP programs in the 1980s
- As set up by people like me
- 1980s up to … ?

Pre-academic EFL/ESL/ESP in the Gulf c. 1985

* What do Ss *actually need* preparing for content study in English?

    (In order)

    - Strong Reading - textbooks and articles => esp. vocabulary
    - Strong Listening - lectures in English => esp. vocabulary, => esp recognition in speech of words known in writing
    - Basic Speaking - for asking questions in class => esp. pronunciation
    - Basic writing - to be graded more for content than correctness

* What did they *get* ? (in order)

    - Grammar focused exercises loaded with unknown vocabulary component
    - Grammar focused writing practice focused on error correction
    - Reading practice with uncontrolled vocab
        - Little distinction between kinds of reading
    - Listening practice with uncontrolled vocab and no instruction beyond practice
    - Random speaking practice with minimal pronunciation focus
        - Ex, little even on p/b and word stress

    (Any guesses why grammar was the major emphasis?)

ESL in the Gulf c. 1985

Based on no particular **learner** analysis

What and how were learners taught in high school? *Unknown*
What are their strengths and weaknesses? *Unknown*
What are their motivations to learn English? *Unknown*

Thus assumptions were simply imported from experience with US/European learners who were normally ~

Acquiring a cognate-rich language
Often for no particular purpose

Holidays
Use heritage language to raise GPA
Relatively low-stakes

ESL in the Gulf c. 1985

Based on no particular **culture** analysis
- Culture normally provides the 'prior knowledge' or 'framing schemata' on which learning can proceed
- The culture assumed in US/UK/European ESL course books often amounts to shopping trips, parties, dating, etc –
  - unfamiliar at best, distracting at worst
- Pedagogical culture often includes task formats that are unfamiliar to the Ss
  - And could be sprung on them in a high-stakes tests
    - 'Unscramble the sequence of sentences summarizing this text'
    - 'Choose the best paraphrase of the main idea of this text'

ESL in the Gulf c. 1985
Based on no particular **culture** analysis

Worse still ~

While imposing foreign schemata, we ignored what could be useful in the schemata the Ss brought with them, such as ~

- A tendency toward cooperative learning and peer teaching
  - Little exploited
  - Viewed mainly as 'cheating'

- Potential processing advantages of a lifetime reading Arabic
  - Greater context-sensitivity (T B Cont'd)

ESL in the Gulf c. 1985

Based on no particular **linguistic** analysis

What affordances and challenges will transfer in from an L1 Arabic? *Unknown*

ESL in the Gulf c. 1985

Based on no particular **research** culture
…… whether international or in-house

- Little or no research among the teaching staff
  - Despite MAs and PhDs
- Not encouraged by institutions through promotion policies
  - Though not particularly discouraged
  - Even where encouraged, was not focused on local realities

Yet, usable research eventually there was (1990s -)

- Either through TESOL Arabia conferences
  - Starting in about 1995
- Or when Gulf veterans went home to do PhDs
  - How long did this research take to get back to the Gulf?

1990s begins to see usable research

- At first observational/anecdotal at conferences
  - Here is a reading passage from Headway (1991) used with low intermediate pre-Commerce Ss in a General English course
    - Presented by me at TESOL-Arabia, Al Ain, UAE, 1995 to show a bad mismatch

The Obs
smatteri
freaks, v
They car
do not a
to be co
the likeli
interest
issue of
buttons
the majo
our lives.

The Observer newspaper recently showed how easy it is, given a suitable story and a smattering of jargon, to obtain information by bluff from police computers. Computer freaks, whose hobby is

1990s begins to se

Another example

- Teacher-talk transcripti
  - These published by Arde

From a chemistry lectur
"Trying to convey the ide
a chemistry lecturer tries
more common analogies
assembly line, all to no a

From a biology lecturer
"The first time I gave a hybridization analogy, I talked about dogs, and then I switched to goats; and then it even dawned on me that some of them aren't going to know that if you mix two different kinds of goats they come out looking in between, and I didn't know all the specific terms there, what their two different breeds of goats are called –

you can talk about mixing colours, but a lot of them don't know their colours yet."

1990s begins to see usable research

Eventually explanatory research emerges

1)   L1 Affordances

Arabic words demand greater contextual input for identification

  Owing to unstated vowels in tri-consonantal roots

  Ex **كتب** **K-T-B = 'books' 'he wrote' 'have somebody write' 'it is written' etc**

  English words are more self-contained, less context determined (Abu-Rabia & Siegel, 1995)

  But imagine English words without vowels: reading, you come to **ct** - is it cat, cut, caught, cot?

Result: Contextual skills developed for Arabic
       should be exploitable

       E.g., in inferring new word meanings from context

       Assuming they know enough of the other words in the context

1990s begins to see usable research
Eventually explanatory

2) L1 Challenges

Arabic speakers problems with word identification and reproduction (spelling) is traced to a different locus of word perception

Reaction Time studies (RT; Randall, 1988)
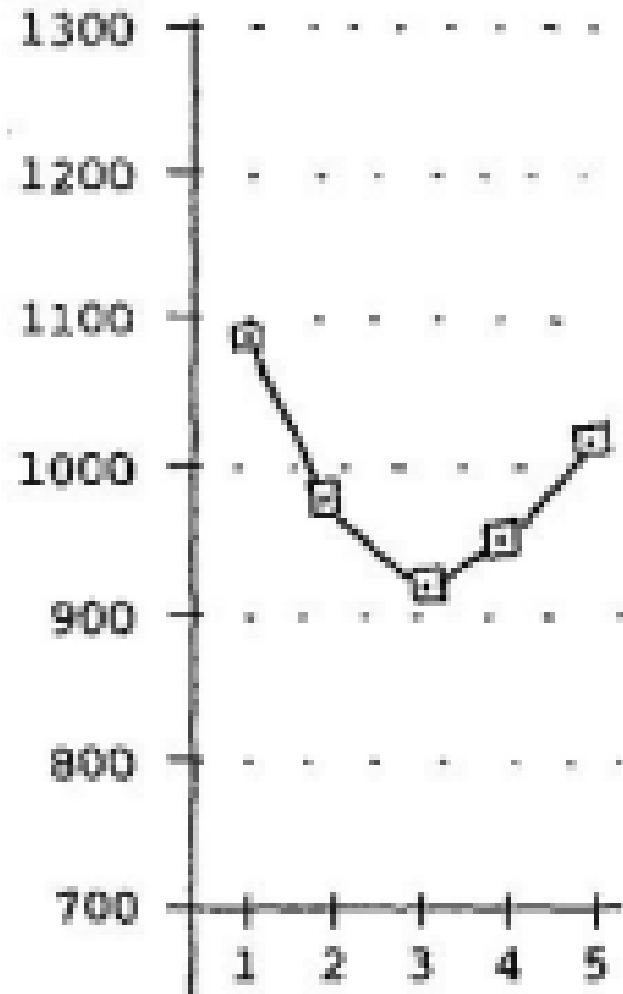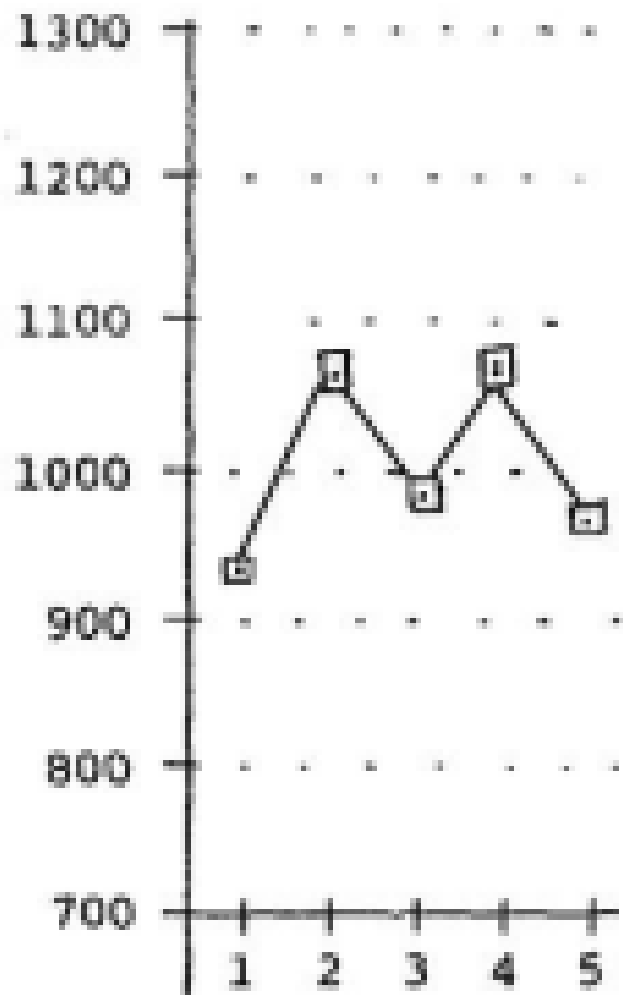
OOOOO OXOOO OOOOX OOXOO OOOOX

"Is there an X in each string?"
Time to answer is measured in msec

Result: Arabic speakers find **x** more quickly when at centre of word

English speakers when **x** is at the left

A 'cognitive process transfer' unlikely to be helpful

But what were the pedagogical implications of these insights?

- How could they be used to improve learning?

One obvious implication:
<< Teach Lexis >>
For both breadth and depth

Most of the 1990s Gulf research focused on *lexis*
Yet I never saw a vocabulary course in 10 years
- No research into the vocabulary Ss knew v. what they needed
  - While 100s of hours were poured into grammar instruction
    - Which is known to be highly interactive with the lexis that carries it
- Yet ironically the major use of class time on any given day was vocab
  - But handled *ad hoc*
    - Without plan, method, system, or explicit testing

So, in general we had set up a fairly under-informed enterprise

Despite the money and number of dedicated and intelligent people involved

- Many with research experience
- That was rarely applied
- Despite the emergence of applied linguistics, 1970s ~
- Despite the amount we didn't know
- Despite the high-stakes risks for the Ss

What might data driven learning contribute to this picture?

Some concrete proposals follow
    Many with prototypes tested in a Gulf context

Discussed via my own software 'www.lextutor.ca'
    Though other software could be used

In two scenarios:
1. A **general** EFL/ESL course
2. A **domain specific** course

Scenario 1. A general EFL/ESL course

Adding a DDL component would need

- a general corpus of the target language
  - Ideally a pedagogically oriented corpus
    - (BNC/Coca, not BNC or Coca)
    - And word lists derived from this corpus
    - Set up to run in a text profiler

- + a corpus of all the Ss' existing learning materials
  - Including old texts and exams if possible

What would having these enable us to do? →

For starters, we could quantify the problem of 'The Observer' text (seen earlier)

This text is readable *with resources* to Ss who know 3,000 word families
Readable *independently* with 7,000 known fams   - Nation (2006)
                                                                            - Laufer (2022) etc

Esp. seen against the ==*vocab testing*== such corpora and lists make possible

Either off-the-shelf k-level tests

LEXTUTOR

MOBILE LEVELS + SIZE TESTS

VLT
Vocab Levels Test

VLT 2

VST
Vocab Size Test

TTV
Test de la taille du vocab

<www.lextutor.ca>

Down Load

Vocab Size Test
Mode = test

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 14

VST - SCORE
TIME (EST):
2022.10.30 / 05.04
LOCATION:
IP_212.221.42.2
NAME: Student

Level	Perc
1k:	40
2k:	0
3k:	0
4k:	0
5k:	0
6k:	0
7k:	0
8k:	0
9k:	0
10k:	0
11k:	0
12k:	0
13k:	0
14k:	0
Size:	400

Recorded here

First 1000 [Go 2]

1. SEE: They **saw** it.
   a. ◯ cut
   b. ◯ waited for
   c. ● looked at
   d. ◯ started

2. TIME: They have a lot of **time**.
   a. ◯ money
   b. ◯ food
   c. ● hours
   d. ◯ friends

3. PERIOD: It was a difficult **period**.

6. DRIVE: He **drives** fast.
   a. ◯ swims
   b. ◯ learns
   c. ◯ throws balls
   d. ● uses a car

7. JUMP: She tried to **jump**
   a. ◯ lie on top of the wa
   b. ● get off the ground suddenly
   c. ◯ stop the car at the of the road
   d. ◯ move very fast

8. SHOE: Where is your sh

Or randomized Yes-No Level Tests
Teacher made, computer assisted

Home > Randomize > RandWords BNC-COCA

BNC-Coca? U.S. COCA (Corpus of Contemporary English) lists blended with U.K. BNC (British Nation

Random item generator v.2

What:
Generate x random items from all 25 BNC-COC lists

How:
Choose list, click *Go* (repeat)

List: 2k heads ∨

Number of Items: 20 ∨ Go »

coca_heads(02).txt

duck
wise
limit
angel
enemy
threat
fail
battery
tower
wipe
wicked
branch
desert
blind
february
salt
reckon
october
flood
argue

1-3k vocabulary knowledge test   PROF xxx / student _____ 5 NOV 22

Put a check mark ✓ *beside each word you know. Some are not real words.*

1k	2k	3k
	angel	
	argue	
	bance	
	battery	
	blind	
	branch	
	degate	
	desert	
	duck	
	eckett	
	eldred	
	enemy	
	fail	
	flood	
	galpin	
	gumer	
	justal	
	limit	
	reckon	
	salt	
	stace	
	threat	
	tower	

Comparing course text profiles to learner test scores showed that…

Text analysis

- The "Observer" text profile is typical of pre-academic ESL course books used in the Gulf
  - Knowledge of 2500-3,000 word-families is the target
    - Higher with domain-specific corpus

Testing

- … while few Gulf Ss were coming out of school with more than 1,000 words
  - Many with less
  - Many Ss' knowledge is giving them ≈ 75% lex coverage in their readings
    - 95% is minimum for basic comprehension
    - Or inference of meaning of novel items

'Observer' with 1,000 words known: 78% coverage

The [1] [____] [2] [____] recently showed how easy it is , given a suitable story and a [3] [____] of [4] [____] to [5] [____] information by [6] [____] from police computers. Computer [7] [____] whose [8] [____] is breaking into [9] [____] systems , don't even need to use the phone. They can [10] [____] their computers [11] [____] with any [12] [____] in the country. Computers do not [13] [____] the [14] [____] issues. But they do [15] [____] the [16] [____] . They allow more [17] [____] to be collected on more [18] [____] of our lives , and [19] [____] both its [20] [____] [21] [____] and the [22] [____] of its [23] [____] [24] [____] from one [25] [____] which might have a [26] [____] interest in it , to another which does not. [27] [____] computer [28] [____] also raise the issue of what is known in the [29] [____] as ' total [30] [____] [31] [____] ' the ability , by pressing a few [32] [____] and waiting as little as a minute , to [33] [____] all the information about us held on all the major government and business computers into an [34] [____] [35] [____] on any [36] [____] of our lives

List of Removed Words (+

agency alter aspect aspects
bluff buttons capabilities
collate connect data (2)
database directly dossier
freaks fundamental hobby
increase instant jargon (2)
legitimate likelihood linkage
modern multiply newspaper
observer obtain official rapid
retrievability risks smattering
transfer unauthorized

35

Similar for a v. basic science text: 75% coverage

Mouse entry or Tabkey + Arrows + Return (shift-Tab to go up)

The [1] [____] 's shape appears to change from a [2] [____] to a half [3] [____] to a whole [4] [____] and back again. These changes are caused by the way sunlight strikes the [5] [____] as it [6] [____] around Earth. Earth's [7] [____] The [8] [____] is a [9] [____] made of rock that [10] [____] around Earth once every Earth days. As it [11] [____] the [12] [____] also [13] [____] once on its [14] [____] in the same amount of time. As a [15] [____] the same side of the [16] [____] the near side always faces Earth. The other side faces away from Earth so you cannot see it from Earth. The [17] [____] does not make its own light [18] [____] is really sunlight [19] [____] from the [20] [____] 's [21] [____] . To [22] [____] means to [23] [____] off. The [24] [____] sunlight makes the side of the [25] [____] facing the Sun look bright Sun. The [26] [____] 's [27] [____] [28] [____] light , but the [29] [____] does not make its own light. If the same side of the [30] [____] always faces Earth , why does the [31] [____] appear to change shape? As it [32] [____] around Earth , the [33] [____] 's near side [34] [____] different amounts of sunlight.

List of Removed Words (+

axis bounce circle (2) crescent
moon (14) moonlight receives
reflect reflected reflecting
reflects result revolves (4)
rotates sphere surface (2)

So any contextual affordance from Arabic will not come into play

36

Following testing and profiling, the first peda-task for DDL is to quickly increase the breadth of general vocabulary knowledge

With a *dedicated vocab course*

- E.g., following a 'Learner as Lexicographer' model
  - After testing, each student is in target=2k or 3k level
  - 1000 words/10 weeks = 100 words per week
  - Some are known – task is build a glossary of those not
    - Ss' decide which words – learner agency etc
    - Find clear example for each in the corpus, + definition, + enter in personal database
  - + weekly quiz

- This is the software I used for such a course, 1990s
- At KSU, then SQU

1) LIST (2k or 3k)
2) CONCORDANCE
3) MEANS TO EXTRACT A GLOSSARY
4) QUIZZES

LIST | A B C D E F G H I J K L

about   above
accept   accident
ache   across
address   admire

129

CONCORDANCE

e and was invited aboard an
on't do something ABOUT that
ently. ¶ L: What ABOUT the
 Mr Harold Thomas ABOUT the
e started talking ABOUT my
 don't open until ABOUT uhm.
on't! ¶ B: Fights ABOUT geti
n the legends are ABOUT the
ell false stories ABOUT the

SOURCE TEX

Q: I'm 16 and I have never been out with
one. My friends have all had lots of girlfri
interested in me. When I ask girls out, th
Now I tell everyone that I have a girl-frie
they believe me. What should I do?

A: People of your age, especially boys, a
experiences. I'm sure some of your frien
shouldn't tell lies, because that will make
people will learn the truth sooner or later
girlfriend. Your time will come.

Word Stacks • 1

⌘Q   WORD STACKS   LAILA

ABOUT

QUIZ: 1->2

عن ، تقريباً

A: People of your age, especially boys, often tell false stories ABOUT their experiences.

Most shops don't open as early as in England, well they don't open until ABOUT uhm… 10 or 11 in the morning.

New Word   Find
Alphabetize   First Wd Today
Throw Out   Print

2 WORDS

Typical 'Learner-as-Lexicographer' Glossary from Oman, 1993

EMPLOY

يستخدم

They EMPLOY a consultant engineer to design it and prepare all the plans.

EMPLOYEE

مستخدم / اجير

The manager began to hire new EMPLOYEES who could put the idea into action.

ENCOURAGE

شجّع

They ENCOURAGED women to break the law.

ENEMY

خصم / عدو

His main ENEMY was the Sheriff of Nottingham, who was always trying to capture Robin but never managed to do it.

ENGAGED

(١) خاطيب / (٢) مشغول

Then the monster attacks and kills not only Frankenstein's friend but also his brother and the woman his brother is ENGAGED to, his brother's fiance Elizabeth.

If you want to phone someone, it's often impossible because they're ENGAGED all the time.

ENGINE

محرك / ماكنة

The ENGINES started to roar, the plane started to shake, and after a bumpy take-off we were suddenly up in the air over the water.

ESCAPE

يفلت من / يهرب

The thieves managed to get in and ESCAPE without setting off the security alarm by cutting off the electricity supply.

With one big surprise:

When I have used this approach, the emergent glossaries become a shared learning resource

Costing reams of paper !

Needed! software to facilitate collaborative dimension

Format of weekly quizzes

Here is "C" week

(Significant labour)

Format for measuring two kinds of word learning.

992 VOCABULARY QUIZ 2 (C)

Part A. Recognition (1 each = 15 points)
Write the number of each word beside the correct definition.

1. candle
2. cabbage   __ conversation
3. candy   __ green vegetable
4. capital   __ teachers use it
5. chalk   __ important city
6. chat

1. centre   __ two people or two things
2. chapter   __ part of a book
3. chimney   __ doctors work there
4. church   __ Christian building
5. clinic
6. couple

1. cry
2. control   __ make unhappy sounds
3. charge   __ make noise with hands
4. cover   __ ask to pay; cost
5. contact   __ have power over
6. clap

1. curious
2. cruel   __ causing pain, hurt
3. crazy   __ not sunny
4. cloudy   __ always asking
5. confident   questions
6. convenient

Part B. Cloze (1 each = 10 points)
Choose from these words to complete the passage.

continued	customers	catch
century	court	congratulations
cheat	clever	caused
connected	curly	contained

Young Computer Criminals

Computers have changed our lives in many ways and made so many things easier to do. Most of us would probably agree that the computer is the most important invention of the 20th (1)............... By now most of us know how to use a computer and these days children learn how to use them at school. This sounds like a great idea. Computer games are good entertainment and educational programs are clearly very useful. But some children are able to do surprising things with computers. For example, in the United States some young computer users found a way to rob banks! These very (2)............ boys made a computer program that wrote thousands and thousands of different telephone numbers. When they (3)............ their program to the bank's telephone, the computer started dialing the numbers. The computer (4)............. dialing the phone numbers, day after day, until the bank's secret telephone number was finally found. Then the boys were able to open the files which (5)........... bank account information, and they started changing things. As you can imagine, you could easily make yourself very rich just by opening a new file and typing some numbers! Of course, bank (6)............. soon noticed that there were problems with their accounts and they started complaining to the bank manager. The police were brought in but it wasn't easy to (7)............... these criminals because no one expected the thieves to be children. In the end the police weren't sure what to do with the boys. They were too young to be sent to a (8)............... of law, and they were certainly too young to go to prison. The bank just had to find a better way to protect its files. Computer criminals have also (9)............... problems for credit card companies in the same way, and these companies have lost a lot of money. Finally, you may be interested to know that students who are good with computers have found a new way to (10)................ Again, by trying thousands of different numbers, they find their way into the computer network of their school and simply change their marks!

Updated software, as used in 2021 by 100+ institutions

Default Corpus
2k Graded Corpu
Default Dictionar
Eng_Arab

[1.] Choose k-list
bnc_heads(3).txt

abroad
absence
accelerate
according
accountant
accuse
ache
adapt
addict
admire
adventure
aerial
aeroplane
affection
aggressive
airline
album
alcohol
alert

to you and offer to talk, which the kids thought was great.

WORD	POS	EXAMPLE	MEANING
ADMIRE	**vb** **2**	You have to ADMIRE the Americans for thinking of such an idea - only they could put together something this big. **3**	**1**

Dictionary Eng_Arab ⌄

< Back   Get

Send words to...

Dictator   Multiconc
ID-words

admire

[7.] Save your corpus findings here
Build a Dictionary of Words, Examples, Meanings (with one click from bottom of Larger Context above). Copy-Paste to an Excel sheet to start building a dictionary of words at your growth-edge.

الإنجليزية-العربية ⌄   ⇄

Type Arabic characters with a virtual keyboard.

flections of 'admire' (v): (⇒ conjugate)
mires   v 3rd person
                  singular
iring   v pres p
hired   v past
hired   v past p

WR Reverse (1)   **WordReference**

ence English-Arabic Dictionary © 2022:

ترجمات رئيسية

الإنجليزي	العربي
admire [sb] ⇒ *vtr* (respect)	يقدّر

admires doctors who nteer to work in third-world countries.

تحترم كريستين الأطباء ال يتطوعون للعمل في بلدان الثالث.

| **admire** [sth] ⇒ (be impressed by) | يُنال |

+ automated weekly tests

(A) Weekly →

Because randomly generated, can be practiced in advance

Advantages of Auto-Randomized Test

- **Cannot** be known in advance ⭐
  - Found + copied etc
    - (All Nation's tests are widely available)
- **Can** be practised ⭐⭐⭐
  - Ss can make any number of practice tests for mobile phone
    - Unlike any fixed test

- **Economical**
  No time-consuming (engagement limiting) labour for teachers ⭐⭐

... + remediated
as needed
with Flashcards

(linked from
test)

(Always with
 TTS)

(B) Mid-Term + Final test →

1. spell from TTS

Also practiceable

Linked from home page

Dictator Spell Words From Speech v.7.2

OPTIONS: Language: en-gb | Gender : female | Voice: Lily| Unit: words | Mode: train | random

TRAINING MODE: Listen to the words, try to spell each, use Guidespell until score = 100%.

Do as Test

Five words from last week .
Practice until you know them then take the test

Click to hear words	WORK SPACE Type the words you hear	GUIDE-SPELL	FEEDBACK SPACE Left to right, this much is correct	OK %
Word 1 >	enemie	Check	enem _ _	80
Word 2 >	employ	Check	employ	100
Word 3 >	encorage	Check	enco _ _ _ _ _	44
Word 4 >	admire	Check	admire	100
Word 5 >	employie	Check	employ _ e	88

Mid-Term + Final test ➔

2. apply meaning to novel context (transfer)

Also practiceable
(E.g. practice online, paper test)
All linked from home page

MultiConc Output

Interactive Quiz Option

Do again (new randomization)

Which word/phrase

admire employ en
Click words for Dictionary E

Questions : 5

History >>

[]

1. ========
    ========
    [001] ADMIRE
    [002] EMPLOY
    [003] EMPLOYEE
    [004] ENCOURAGE
    [005] ENEMY
    [006] t walked o
    [007] sorry for

2. ENCOURAGE

    [001] iters but
    [002] a step fu
    [003] d over the
    [004] ritish goo

Multiple Concordances (Corpus=*corpus_graded_2k.txt*)

| EMPLOY | EMPLOYEE | ADMIRE |

Hits: 007

```
[001]  businessmen rarely wanted to  _____  paid workers. The shipping bu
[002]  a bridge or a hospital. They   _____  a consultant engineer to desi
[003] ge of cheap communications to  _____  data-entry clerks in Barbados
[004] n for  example, London, might  _____  clerks in Belfast, or a Paris
[005] are of t he abilities she can  _____  to overcome this handicap. Li
[006] fter the inquiry. Nobody will  _____  me as an officer again. I hav
[007]  businessmen rarely wanted to  _____  paid workers. The shipping bu
```

Hits: 007

```
[001] g. They just work. A contract  _____  can charge double the cost of
[002]  double the cost of a regular  _____  and still be a bargain.  FOR
[003] ug Enforcement Administration  _____  won $180,000 in Probst v. Ren
[004] Housing and Urban Development  _____  , in Barvick v. Cisneros et al
[005]  it 'The Spirit of Delta.' As  _____  morale crashed, customer comp
[006] oday that it had dismissed an  _____  for unauthorized options trad
[007] urely the  person is still an  _____  ? LD: Well yes, you may think
```

Hits: 007

```
[001] Salisbury. Maudsley seemed to  _____  the name Court Place. He prob
[002] th my bicycle. Everyone would  _____  it. I had not learned to ride
[003] hen suddenly kind. He did not  _____  me particularly, and I felt t
[004] t. It's enormous! You have to  _____  the  Americans for thinking o
[005] dworking. I think you have to  _____  them for that. The hard  work
[006] , you know very well that you  _____  him.' Oh, Aunt Penniman!' sai
[007] herine again. Perhaps she did  _____  him though this did not seem
```

Summary: Vocab course

Words are ~

- Identified by testing

- Numerous

- Met in a variety of aspects, contexts, and tasks
  - In and out of context

- Will be re-met *in situ* in regular learning materials (the corpus)

+ but DDL can also support other areas of a language course →

Corpus based grammar

1 prep

tutorial→

Grammar intuition v. corpus data for error correction

Prepositions | Singular-Plural | Word Order | Gerund-Infinitive | Simple Past-Present Perfect | Conditionals | Formulaics | French

A. Prepositions

Num	Error sentence	Data	Correction space	CHECK	Help	FB
1.	He's going to home.	CONC	He's going to home.	Check	Help	
2.	The people are fed up of so many elections.	CONC	The people are fed up of so many elections.	Check	Help	
3.	Tell to me everything she said.	CONC	Tell to me everything she said.	Check	Help	
4.	They listen the radio every morning.	CONC	They listen the radio every morning.	Check	Help	
5.	(MULTI) She is interesting to watching birds.	CONC	She is interesting to watching birds.	Check	Help	

B. Singular-Plural

Home > Concordancers > English Input > Output   (« Back keeps original settings)   Copiable extract-link

Concordance for **home** in **brown_bncw.txt** ▮ Dictionary/TTS

MODIFY:

equals ▼ ; home   Brown_BNC Writ. (2+m)>   SORT ▼   key ▼ +ASSOC go   ON L|R ▼ FOR

25,000 ▼ LINES @ 80 ▼ WIDE **Go**

027. ☐ got to his feet. He went back to each HOME place, turned and walked to the day
028. ☐ hing for you to do", Eddie said, "is go HOME. You don't belong in professional b
029. ☐ said Samuel Burns, "probably went right HOME and poured himself into a boiling b
030. ☐ d no one could fill. The Brainards went HOME early. Martin realized, later on, t
031. ☐ be no trouble tonight so that I can go HOME early. Unified state more longing t
032. ☐ n town Tegucigalpa, and it's time to go HOME. And here, at last, come the Kids f
033. ☐ ght in Tegucigalpa, and it's time to go HOME. And here, at last, come the Kids f
034. ☐ ercenary, Mr Bob Denard, and his men go HOME. -- AP. Rebels attacked. Planes bom

When I had finished I went to home. [Revision 1] The next day I got up...

Corpus based grammar

1 as prep tutorial

2 as embedded links in Ss writing →

Gaskell & Cobb 2004

http://www.lextutor.ca/scripts/cgi-bin/wwwassocwords.exe?Corpus=Brown.txt&SearchStr=home&AssocWord=go&Associate=Left&D1=6&Maximum=21

```
1    . But if you're worried you can go home and check"-      "I can't leave the
2    e left, it was dark and time to go home and cook supper for her husband. "
3    ds who seemed determined not to go home at all. Only a plea from the house
4     to make when they are about to go home, but drinking is their sickness. Y
5    row some money from someone and go home by bus? I could send the money rig
6     saying "American imperialists, go home". Chin up, Soapy. @  B01 0960  8
7     miles for the burial would not go home, for she might die any time; but t
8    ".      "Aren't you ever going to go home"?      "It sure as hell doesn't look
9    t without flinching.     "I said go home, Joseph. You've got no business up
10   near the Dutch Reformed.     "Go on home"! Ludie screeched at him. "Someone
11   or, she would tell him, "If you go home now, I'll scream". More often than
12   may go up near enough to hurt your home or to hurt its value.  F13 1330  4
13    are you now"? or an "I want to go home", or whatever- but a nonverbal one
14   een. The little Ito girl had to go home. She has a pretty bad cold".     "I
15   o beat people ... I want ... to go home".     These were the last words he
16   thirty minutes after the others go home this afternoon and work your probl
17   Goddamn. Goddammit. Just let me go home to Jersey, back to the shore, oh,
18   ly no place to go, finally, except home to Mae.     At the gate he slowed,
19   r said, was let manual laborers go home Tuesday night for some rest. Work
```

0

Morpho-logy in context

The Observer newspaper recently showed how easy it is, given a suitable story and a smattering of jargon, to obtain information by bluff from police computers. Computer freaks, whose hobby is breaking into [1] [_____ ▾] systems, don't even need to use [2] phone.

They can connect their computers [3] [_____ ▾] with any database in the country. Cor do not alter the fundamental issues. But [4] [_____ ▾] do multiply the risks. They allow [5] [_____ ▾]

data to be collected on more [6] [_____ ▾] of our lives, and increase both [7] [__ ▾] ra

aspect
aspected
aspects

retrievability and the likelihood authorized transfer from one agency whi

might have a legitimate interes [__ ▾] another which does not. Modern com

capabilities also raise the issue of what [10] [_____ ▾] known in the jargon as 'total [11] linkage' the

Reading with resources

MEDIA SPACE

Save on Lextutor as recycled_plastic[6].html

HYPERTEXT FILE: recycled_plastic[6]
Dico: Eng_Eng | **TTS accent:** US | **Platform:** computer

Select a word (double-click) & then click DEFINE ⬇
or word/paragraph (triple-click) & SPEAK ⬇

WORD BOX   SELECT

landfills

Words go to...   Dictator   MultiConc   ID-Words   DEFINE   SPEAK

◉ Normal
○ Slow
○ Stop

Search   English definiti

English © 2022

land•fill /ˈlænd ˌfɪl/
n.

1. a low area of land [countable]
   built up from deposits of soli
   layers and covered by soil.

2. the solid [uncountable]
   garbage itself.

VOA News in Regular English -

Places like these are where most of the world's plastics end up - in landfills. And a fair amount of it ends up in the world's oceans, but a new British recycling firm wants it to end up here.

There is a great demand right now for recycled material; there's a great deal of interest in getting recycling into the plastics industry, especially, and so we are really going to facilitate the ability to get plastic into the circular economy.

The firm called Recycling Technologies is breaking down plastic then turning it into fuel oils, or feedstock, to make

Reading with resources + listening

And the U-Shaped curve?
(Of Arabophones' word recognition)
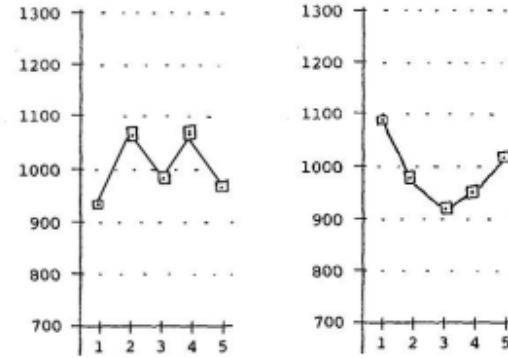
No simple solution to this

- But ***massive exposure*** and ***fluency practice*** are a first guess at a pedagogy
  - Corpus is a massive source of input
    - If it can be carved up in an engaging manner

  - Fluency practice →

Fluency practice

- This can not happen with The Observer text
  - It is the 'hard going' type of reading
  - Which has some uses but not a steady diet
- Fluent reading takes place with texts that contain ~
  - 98% well-known words
    - = recognized in < 900 msecs
- Where do such texts come from
  - Simplified stories, extensive reading – rare in 1990s Gulf but **always** successful when used
    - Bookworms, SRA Kits
- Or make your own
  - **Computer-assisted text modification →**

(1) Achieve 98% known words with text modification

The Observer newspaper recently showed how easy it is, given a suitable story and a smattering of jargon, to obtain information by bluff from police computers. Computer freaks, whose hobby is breaking into official systems, don't even need to use the phone. They can connect their computers directly with any database in the country. Computers do not alter the fundamental issues. But they do multiply the risks. They allow more data to be collected on more aspects of our lives, and increase both its rapid retrievability and the likelihood of its unauthorized transfer from one agency which might have a legitimate interest in it, to another which does not. Modern computer capabilities also raise the issue of what is known in the jargon as 'total data linkage' the ability, by pressing a few buttons and waiting as little as a minute, to collate all the information about us held on all the major government and business computers into an instant dossier on any aspect of our lives

Profile summary		
K	#	cumul%
K-1	132	78.6
K-2	16	88.1
K-3	11	94.6
K-4	1	95.2
K-5		
K-6	2	96.4
K-7	2	97.6
K-8		
K-9		
K-10		
K-11	2	98.8
K-12	1	99.4
K-13		

the-1 observer-2 newspaper-2 recently-1 showed-1 how-1 easy-1 it-1 is-1 given-1 a-1 suitable-1 story-1 and-1 a-1 smattering-12 of-1 jargon-7 to-1 obtain-3 information-1 by-1 bluff-6 from-1 police-1 computers-1 computer-1 freaks-6 whose-1 hobby-2 is-1 breaking-1 into-1 official-2 systems-1 do-1 not-1 even-1 need-1 to-1 use-1 the-1 phone-1 they-1 can-1 connect-2 their-1 computers-1 directly-2 with-1 any-1 database-3 in-1 the-1 country-1 computers-1 do-1 not-1 alter-2 the-1 fundamental-3 issues-1 but-1 they-1 do-1 multiply-4 the-1 risks-2 they-1 allow-1 more-1 data-3 to-1 be-1 collected-1 on-1 more-1 aspects-3 of-1 our-1 lives-1 and-1 increase-2 both-1 its-1 rapid-2 and-1 the-1 likelihood-2 of-1 its-1 unauthorized-3 transfer-2 from-1 one-1 agency-3 which-1 might-1 have-1 a-1 legitimate-3 interest-1 in-1 it-1 to-1 another-1 which-1 does-1 not-1 modern-2 computer-1 capabilities-2 also-1 raise-1 the-1 issue-1 of-1 what-1 is-1 known-1 in-1 the-1 jargon-7 as-1 total-1 data-3 linkage-3 the-1 ability-1 by-1 pressing-1 a-1 few-1 buttons-2 and-1 waiting-1 as-1 little-1 as-1 a-1 minute-1 to-1 collate-11 all-1 the-1 information-1 about-1 us-1 held-1 on-1 all-1 the-1 major-1 government-1 and-1 business-1 computers-1 into-1 an-1 instant-2 dossier-11 on-1 any-1 aspect-3 of-1 our-1 lives-1

The Observer newspaper recently showed how easy it is, given a suitable story and a BIT of POLICE LANGUAGE, to GET information from police computers. Computer CRIMINELS, whose PLEASURE is breaking into official systems, don't even need to use the phone. They can connect their computers directly with any data SOURCE in the country. Computers do not CHANGE the BASIC issues. But they do INCREASE the risks. They allow more data to be collected on more PARTS of our lives, and increase both its ACCESS and the likelihood of its ILLEGAL transfer from one OFFICE which might have a REASONABLE interest in it, to another which does not. TODAY'S computer ABILITIES also raise the issue of what is known as 'total data linkING' - the ability, by pressing a few buttons and waiting as little as a minute, to BRING TOGETHER all the information about us held on all the major government and business computers into a FILE on any PART of our lives

Edit-to-a-Profile

Edit, Check with

RE-VP

Profile summary		
K	#	cumul%
K-1	143	87.2
K-2	16	97.0
K-3	5	100.0
K-4		

the-1 observer-2 newspaper-2 recently-1 showed-1 how-1 easy-1 it-1 is-1 given-1 a-1 suitable-1 story-1 and-1 a-1 bit-1 of-1 police-1 language-2 to-1 get-1 information-1 from-1 police-1 computers-1 computer-1 criminels-1 whose-1 pleasure-2 is-1 breaking-1 into-1 official-2 systems-1 do-1 not-1 even-1 need-1 to-1 use-1 the-1 phone-1 they-1 can-1 connect-2 their-1 computers-1 directly-2 with-1 any-1 data-3 source-3 in-1 the-1 country-1 computers-1 do-1 not-1 change-1 the-1 basic-1 issues-1 but-1 they-1 do-1 increase-2 the-1 risks-2 they-1 allow-1 more-1 data-3 to-1 be-1 collected-1 on-1 more-1 parts-1 of-1 our-1 lives-1 and-1 increase-2 both-1 its-1 access-2 and-1 the-1 likelihood-2 of-1 its-1 illegal-2 transfer-2 from-1 one-1 office-1 which-1 might-1 have-1 a-1 reasonable-1 interest-1 in-1 it-1 to-1 another-1 which-1 does-1 not-1 today-1 computer-1 abilities-1 also-1 raise-1 the-1 issue-1 of-1 what-1 is-1 known-1 as-1 total-1 data-3 linking-3 the-1 ability-1 by-1 pressing-1 a-1 few-1 buttons-2 and-1 waiting-1 as-1 little-1 as-1 a-1 minute-1 to-1 bring-1 together-1 all-1 the-1 information-1 about-1 us-1 held-1 on-1 all-1 the-1 major-1 government-1 and-1 business-1 computers-1 into-1 a-1 file-2 on-1 any-1 part-1 of-1 our-1 lives-1

(2) Reduce word recognition speed
Get it closer to native speaker's 700-900 msecs. How?

RT for word recognition can improve with practice
- A certain amount of practice happens with fluent reading itself
  - E.g. of modified texts etc
- Not quickly

But RT can also be tackled directly
- Speed-rewarding computer games can drop RT
  - 'Word Coach' study (Cobb & Horst, 2011) found a 43% decrease in recognition time for common words after a few hours game use

Explore

Pre-Med English, KSU, 1986

But off-the-shelf games have problems
- No corpus, dictionary based
- Unadaptable to user content
- Extra hardware $$$

Solution

Adapt an RT research tool that runs over WWW on mobile phone

Using common RT designs - such as lexical decision task

This one asks 'Are these real words?'

Records are kept when errors=0

Solution
Adapt an RT research tool that runs on mobile phone

Or, using a design from the U-shape curve study:

**"Are these
 words same or different?"** →

From Ryan & Meara
(1991)
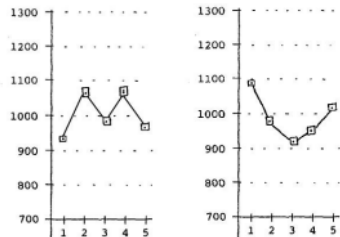
Results

```
NAME: tom ROUTINE: same-diff_1
Mean RT(secs)=1.47
Std_Dev=0.211
Errors=2 (16.67%)
+/- mean

(To put in Excel select 5 columns and paste)

Fri Oct 21 2022 17:12:21 GMT-0400 (Eastern Daylight Ti

GRP       ANS       ITEM           OK?       RT        +/-Mean

YES       Yes       sufficient+sufficient    ok        1.543
YES       Yes       revolution+revolution    ok        1.401
YES       Yes       photograph+photograph    ok        1.005
YES       Yes       management+management    ok        1.461
YES       Yes       distribute+distribute    ok        1.407
YES       Yes       department+department    ok        1.474
NO        Yes       photograph+photogrph     x         1.825
NO        Yes       experiment+expriment     x         1.537
NO        No        sufficient+sufficint     ok        1.354
NO        No        revolution+revoltion     ok        1.509
NO        No        management+managment     ok        1.338
NO        No        automobile+autmobile     ok        1.781
```

And finally
DDL and responsible testing

Testing is about knowledge transfer to a novel context

- As noted, corpus is particularly transfer friendly
  - Provides numerous opportunities to practise transfer
    - RE-randomizations from the corpus

- Test must require only transferable knowledge
    - Transferrable from what was taught
    - How can this be guaranteed?
  1. Familiar task types
  2. Familiar vocabulary and grammar

Let's look just at vocab:
How can a test be tested to contain only words previously met? →

61

Software for ethical testing: Text-Lex Comparison

- Course Corpus is entered as the 'first' text
  - Exam Draft as 'second' text
- Output is percent of lexical overlap
  - Shared words / unique words x 100
- Since < 5% novel items can be inferred, minimum overlap is 95%
  - If inference had been part of the teaching
  - Therefore re-write until this is the case

Example from a recent in-house reading course in Canada →
  - KSU-1980s must have been similar

Home > TL Compare   New - better list comparison (Oct 2021)

First text: COURSE CORPUS (1504 families)

Second tex

Home >> Text_Lex_Compare Input via 'back' (to keep inputs) or **Direct >> Output**

EDITOR RAISE / LOWER RECYCLING INDEX (OR GET VISUAL FOR OWN WORDLIST IN A TEXT)   ↑Back

Works best with smallish second/last text

Editor is easier to use without stoplists - NO STOPLISTS RUNNING

Token Recycling Index : 2464 shared / 7413 total = 33.24%

Token Coverage : 123 from 1st in 2nd / 168 total in 2nd = 73.21%

Types Recycling Index : 69 shared / 2137 total = 3.23%

Types Coverage : 69 from 1st in 2nd / 110 total in 2nd = 62.73%

EDITING FEATURE
Here is the complete 2nd text with **unshared items** underlined - EDIT THE 2nd TEXT to raise the recycling of items in previous text(s): by, where possible, removing or replacing new words with words that have already appeared in previous text(s) -- as reviewed (in families to save space) in the textareas on left. Click Recycle to view your progress

TOKE
FAMI

.88%
) = 75.64%

TYPES UNSHARED (=UNIQUE TO FIRST TEXT/S)

1958 words

,pde,bit,adippm able aboriginal abroad
abundant abundantly accept access
accessed accessible accomplished
according account achievement across act
action actively activism activities acts
actually added addiction addresses
adequately administers ads adulthood

Integrate
←appropriate
forms of
word
families)
from lists
on left

2ND/LAST TEXT WITH **UNSHARED** TYPES
🔄 **Recycle**
underlined
To raise recycling, take out underlined words
or rewrite with used words & Recycle

The Observer newspaper recently
showed how easy it is, given a
suitable story and a smattering of
jargon, to obtain information by
bluff from police computers.
Computer freaks, whose hobby is

63

Of course, ideally DDL work moves towards more and more *formative evaluation*

Ideally in collaborative setting

- Not mass summative testing, which better suits the rules-based grammar class

End Scenario 1

General ESL/EFL course

Scenario 2:
DDL for a domain-based ESL or ESP course

All of the foregoing,
- Vocab testing
- Text modification
- TTS work + Corpus grammar
- Fluency + intensive reading
- Ethical testing…

- *PLUS* a new corpus: domain texts divided by themes or units

Domain corpus will pose new vocab challenges

- Need for an explicit *list*
  - For 1. testing and 2. vocab supplement alongside reading

- However, making a list ≠ borrowing Nation's list
  - List cannot be based only on frequency
    - High frequency item can be all from one text or course unit
    - We need to find word families that are both frequent plus occur across texts

- For this we need a software package called *Range*

Example –
'medical' corpus of Dr House TV series, all 8 seasons

- Input to Range as a zip file

Range output looks like this →
With 1+2k stopped
+ proper nouns stopped

See ⇩ bottom⇩ for filenames, stoplists, offlist, profile, output filtering (freq x range), & Excel-copiable version of output at ⇩ bottom⇩

TITLE: house.zip **Post-Analysis=>** | VP (token coverage) | | Range Profile |

| INPUT FILES: 8 | FAMS: 4008 | CLASSIFIABLE TOKENS: 5,127 |STOPLISTS: k1_bncoca k2_bncoca + Proper nou |

Click headings to sort ⬍   (In the table, **T1(25)** means text T1 has 25 occurrences of the word, etc)

000.	Fams	Freq	Range	VP	T1	T2	T3	T4	T5	T6	T7	T8
101.	lung	123	8	4	T1(1)	T2(14)	T3(16)	T4(9)	T5(7)	T6(4)	T7(30)	T8(42)
111.	cancer	109	6	3	T1(2)	T2(20)	T3(14)	T4(6)	T5(59)			T8(8)
168.	symptom	64	8	3	T1(11)	T2(8)	T3(14)	T4(4)	T5(9)	T6(5)	T7(6)	T8(7)
173.	tumour	63	7	4	T1(16)	T2(24)	T3(4)	T4(2)	T5(7)	T6(3)		T8(7)
263.	hallucinate	38	7	8	T1(1)	T2(15)	T3(8)	T4(2)	T5(7)	T6(4)	T7(1)	
289.	confirm	34	7	3	T1(2)	T2(2)	T3(3)	T4(9)	T5(2)		T7(13)	T8(3)
309.	diagnosis	31	8	4	T1(15)	T2(3)	T3(1)	T4(1)	T5(4)	T6(1)	T7(2)	T8(4)
317.	scan	31	7	3	T1(4)	T2(3)	T3(12)	T4(5)	T5(5)	T6(1)		T8(1)
331.	patients	30	7	3	T1(10)	T2(2)	T3(2)		T5(3)	T6(5)	T7(6)	T8(2)
347.	transplant	28	5	4		T2(2)	T3(2)		T5(10)		T7(7)	T8(7)
405.	fluid	23	7	4	T1(3)	T2(4)	T3(8)		T5(2)	T6(2)	T7(1)	T8(3)
418.	pill	22	8	4	T1(3)	T2(1)	T3(2)	T4(6)	T5(1)	T6(6)	T7(1)	T8(2)
437.	theory	21	6	3	T1(1)	T2(3)	T3(2)	T4(5)	T5(7)			T8(3)
443.	neurosurgery	20	2	12		T2(2)					T7(18)	
452.	allergy	19	4	5	T1(2)			T4(12)		T6(1)		T8(4)
468.	terror												
477.	negative											T7(2)	T8(5)
493.	function											T7(4)	
498.	seizure											T7(9)	
505.	collapse												T8(4)
514.	steroid											T7(2)	T8(1)
516.	abnormal												T8(1)
522.	contrast												
530.	disorder	15	4	3	T1(1)		T3(11)				T7(2)	T8(1)
542.	surgeon	15	6	4		T2(8)	T3(2)	T4(2)	T5(1)		T7(1)	T8(1)

OUTPUT FILTER

Extract data by Frequency >= | 8 ˅ | and Range >= | 4˅ |
Format: [⦿ As above ◯ Words only]   `Go »`
Build Frequency- and Range-based word lists; check corpus coverage a

FAM		FREQ	RANGE
110.	cancer	109	6
167.	symptom	64	8
172.	tumour	63	7
262.	hallucinate	38	7
288.	confirm	34	7
308.	diagnosis	31	8
316.	scan	31	7
330.	patients	30	7
346.	transplant	28	5
404.	fluid	23	7
417.	pill	22	8
436.	theory	21	6
451.	allergy	19	4
476.	negative	18	6
492.	function	17	6
497.	seizure	17	6
513.	steroid	16	5
515.	abnormal	15	6
521.	contrast	15	4
529.	disorder	15	4
541.	surgeon	15	6
546.	urine	15	7
554.	biopsy	14	7
574.	procedure	14	6
582.	abdomen	13	4
584.	antibiotic	13	6
593.	cane	13	5
594.	chart	13	4
601.	fake	13	5
606.	inflame	13	4
607.	lesion	13	4
616.	sane	13	6

631.	consent	12	4
644.	immune	12	6
645.	jerk	12	5
648.	medication	12	5
650.	metaphor	12	5
655.	proof	12	6
656.	rape	12	4
659.	solve	12	5
665.	approve	11	5
667.	arrhythmic	11	4
674.	consult	11	5
678.	diagnose	11	5
679.	differential	11	4
682.	false	11	4
686.	joint	11	5
688.	lobe	11	4
693.	miserable	11	6
698.	pulmonary	11	6
709.	structure	11	4
711.	suicide	11	5
712.	tissue	11	4
717.	abuse	10	6
720.	analyse	10	4
733.	delude	10	4
735.	elevate	10	4
739.	failure	10	5
750.	radiate	10	4
752.	response	10	5
760.	therapy	10	4
769.	cardiac	9	4
776.	code	9	4
777.	coincide	9	4
802.	pants	9	5
808.	psychiatry	9	4
809.	pulse	9	6

811.	relevant	9	4	3
818.	severe	9	5	3
826.	sweat	9	4	3
827.	toxin	9	6	6
837.	assign	8	5	3
838.	autopsy	8	4	8
862.	focus	8	5	3
863.	genetic	8	5	3
871.	irony	8	6	3
890.	respond	8	4	3
902.	technical	8	5	3

JUST < 100 WORDS PRE-MED Ss ARE GOING TO MEET OVER AND OVER IN THESE MATERIALS

ADJUST PARAMS FOR LARGER OR SMALLER

NOW WHAT IS NEEDED IS A PEDAGOGY TO DELIVER THESE WORDS

IN A VOCAB FOCUS COURSE TO ACCOMPANY THE USUAL READING ACTIVITIES

TWO STEPS

	A	B	C	D	E	F	
73	837	assign	8	5	3	T1(1)	T2(
74	838	autopsy	8	4	8		T2(
75	862	focus	8	5	3	T1(2)	
76	863	genetic	8	5	3	T1(3)	T2(
77	871	irony	8	6	3	T1(1)	T2(
78	890	respond	8	4	3	T1(2)	
79	902	technical	8	5	3		T2(
80			FREQ	RANGE	VP		
81							
82					4.03	AV	
83					1.63	SD	
84							

- **FIRST** Are Ss ready for the specialist list?

- Average k-level of Dr House list = 4
  - SD=1.63
    - Thus k=2 to k=6
  - Test Ss with a Levels type test as before

- Ss who know fewer than 3,000 words of General English will need remediation to reach 3k
  - Use procedures already described
- REF Hwang & Nation

SECOND
When ready, Ss build a glossary of the Core Lexicon (about four weeks)
- Then incorporate the rest from their reading

How?

Using these DDL tools
(1) the specialist corpus and
(2) an ***Interactive Database***
- *... collaboratively*

(Recall 💡- the need for collaborative software for the 'learner as lexicographer' project)

Here it is →

EMPLOY

يستخدم

They EMPLOY a consultant engineer to design it and prepare all the plans.

EMPLOYEE

مستخدم / اجير

The manager began to hire new EMPLOYEES who could put the idea into action.

ENCOURAGE

شـجّع

They ENCOURAGED women to break the law.

ENEMY

خصم / عدو

His main ENEMY was the Sheriff of Nottingham, who was always trying to capture Robin but never managed to do it.

ENGAGED

(١) خاطيب / (٢) مشغول

Then the monster attacks and kills not only Frankenstein's friend but also his brother and the woman his brother is ENGAGED to, his brother's fiance Elizabeth.

If you want to phone someone, it's often impossible because they're ENGAGED all the time.

ENGINE

...ك / ماكنة

The ENGINES started to roar, the pl started to shake, and after a bumpy off we were suddenly up in the air o the water.

ESCAPE

لت من / يهرب

The thieves managed to get in and ES without setting off the security alar cutting off the electricity supply.

ALL ENTRIES: Quiz Options>> [⦿ Print | ⦿ Interactive]

Check> | 1st 10 | 2nd 10 | 3rd 10 | 4th 10 | 5th 10 | 6th 10 | 7th 10 | 8th 10 | 9th 10 | 10th 10 | | Any 20 | None | Quiz checked

v.8.4

V. GROUP
+ LEX
+ Demo (en)

OS=Win
BROWSER=Chr
PHP 7.3.33

4	☐	**virus**	3	They are trying to stop the spread of the flu **virus**.	N	A very small living thing that causes infectious illnesses	Arts	Giovanna
5	☐	**drug**	1	The big **drug** companies make huge profits.	N	a medicine, or a substance for making medicines	Arts	Giovanna
6	☐	**penicillin**	9	Since the discovery of **penicillin**, the treatment of syphilis has been revolutionized.	N	a type of medicine that is used to treat infections caused by bacteria	Arts	Giovanna
7	☐	**antibiotic**	5	Thirteen strains were resistant to three or more **antibiotic**s.	N	a drug that is used to kill bacteria and cure infections.	Arts	Giovanna
8	☐	**vaccine**	5	Doctors worried that there would not be enough **vaccine** for everyone who needed it.	N	a substance which contains a weak form of the bacteria or virus that causes a disease and is used to protect people from that disease	Arts	Giovanna
9	☐	**fall**	1	leaves of any colors were all over the place during **fall**	N	the season of the year that comes after summer and before winter -	Comm	faby1
10	☐	**résumé**		she handed over her **résumé** and waited.	N	a brief written account of educational and professional qualifications and experience	Comm	faby1
11	☐	**recess**	6	let's have a short **recess** between classes	N	a temporary withdrawal from or stopping of the usual work or activity - - a break.	Comm	faby1
12	☐	**downtown**		let's go **downtown** to do some shopping	N	of, pertaining to, or situated in the main business section of a city. -	Comm	faby1
13	☐	**atm**		if you need cash you can draw it from an **ATM**	N	automated-teller machine.	Comm	faby1
14	☐	**highway**	3	from a big town to another better get the **highway** to avoid local traffic	N	a main road, esp. one between towns or cities.	Comm	faby1
15	☐	**fall**	1	leaves **fall** during **fall**	N	third season in a year	Comm	faby1
	☐		1	She reflects her personality with	V	to have on the body as clothing,	Arts	Alby

V.Mobile pad/phone

Go v.French

COMPUTER
See all
2372 Words

Add Entry

Add Stude...

Edit Entries

Need Your Own Gp Lex? Write Tom

Get list+VP

Any research on this?

Speak the selection

ttps:/.../tutor.ca/group_lex/d...o/index.php?tri=date&ling=en

74

Lexical Entry Form

	NEW WORD	EXAMPLE	WORD CLASS	DEFINITION	GR
0	**symptom**	He had every symptom you'd expect of as drug addict	NOUN ⌄	sign of disease عرَض (لمرض)	Arts
	Max 25 characters	Max 140 chars., 89 left		Max 140 chars., 113 left	

Dictionary for word in first box >>

Choose from nine: Eng_Arab ⌄   Get

WordRefe

الإنجليزية-العربية

Type Arabic characters with a virtual keyboard.

symptom

🔊)) LISTEN: UK

(simp'təm), /ˈsɪmptəm/ :UK:* /ˈ

WR Reverse (1)   W

WordReference English-Arabic Dictionary © 2022:

ترجمات رئيسية

العربية

(sign of disease)   عرَض (لمرض)

Be sure to describe all of your symptoms to the doctor.

احرص أن تصف للطبيب كل أعراضك.

(indicator)

ALL ENTRIES: Quiz Options>> [⚪ Print | 🔵 Interactive]

Check> | 1ˢᵗ 10 | 2ⁿᵈ 10 | 3ʳᵈ 10 | 4ᵗʰ 10 | 5ᵗʰ 10 | 6ᵗʰ 10 | 7ᵗʰ 10 | 8ᵗʰ 10 | 9ᵗʰ 10 | 10ᵗʰ 10 | | Any 20 | None | **Quiz checked** | **Thanks musta_k**

v.8.4

v. GROUP
+ LEX
+ Demo (en)

OS=Win
BROWSER=Chr
PHP 7.3.33

V.Mobile
pad/phone

Go v.French

COMPUTER

#	Qz	WORD ⇕	VP ⇕	EXAMPLE	POS ⇕	DEFINITION	GRP ⇕	STUDENT NAME ⇕	TIME ⇕
1	☐	**symptom**	3	He had every **symptom** you'd expect of a drug addict	N	sign of disease - (لمرض) | عرَض	Arts	musta_k	2022.10.27 11:34
2	☐	**trip**	1	The boat **trip** down the Amazon was great.	N	an occasion when you go somewhere and come back again	Arts	faby1	2022.10.06 12:04
3	☐	**cantanke...**	14	a **cantankerous** mule	ADJ	difficult or irritating to deal with	Comm	frauclaud..	2022.07.11 10:25
4	☐	**automati...**	5	• Some say **automation** is the main culprit	N	the use of computers and machines instead of people to do a job	Arts	Frauanto	2022.07.11 10:24
5	☐	**virus**	3	They are trying to stop the spread of the flu **virus**.	N	A very small living thing that causes infectious illnesses	Arts	Giovanna..	2022.07.11 07:11
6	☐	**drug**	1	The big **drug** companies make huge profits.	N	a medicine, or a substance for making medicines	Arts	Giovanna..	2022.07.11 07:09

76

Or by mobile

… Ss practice for weekly quiz with this

QUIZ 1 - contexts from AA GP LEX 27 Oct 22   Quiz 2 -*n*

< ALL AA GP LEX

Check   Questions: 6   OK: 0   Tries: 0   %: 0

	NEW WORD	EXAMPLE	WORD CLASS	D
1		Thirteen strains were resistant to three or more ___ s.	N	a drug that is used to kill bacteria
2		Since the discovery of ___ , the treatment of syphilis has been revolutionized.	N	a type of medicine that is used to
3		He had every ___ you'd expect of a drug addict	N	sign of disease - (للمرض) | عَرَض
4		Doctors worrie ___ e would not be enough for ___ who needed it.	N	a substance which contains a wea a disease and is used to protect p
5		The big ___ anies make huge profits.	N	a medicine, or a substance for ma
6		They are trying ___ e spread of the flu ___ .	N	A very small living thing that caus

antibiotic
drug
penicillin
symptom
vaccine
virus

... then the weekly or final paper quiz is either a variant of what they have practised ('near transfer')

	NEW WORD	Fill the gap in the example	WORD CLASS	DEFINITION
1		Doctors worried that there would not be enough _____ for everyone who needed it.	N	a substance which contains a weak form of the bacteria or virus that causes a disease and is used to protect people from that disease
2		He had every _____ you'd expect of a drug addict	N	sign of disease - (المرض)

	NEW WORD	Supply the defined word	WORD CLASS	DEFINITION
1			N	a substance which contains a weak form of the bacteria or virus that causes a disease and is used to protect people from that disease
2			N	sign of disease - عَرَض | (المرض)

	NEW WORD	Match word to definition	WORD CLASS	DEFINITION
1		antibiotic	N	a substance which contains a weak form of the bacteria or virus that causes a disease and is used to protect people from that disease
2		bacterium	N	sign of disease - عرَض | (لمرض)
3		drug	N	A very small living thing that causes infectious illnesses
		penicillin		
4		symptom	N	a type of medicine that is used to treat infections caused by bacteria
5		virus	N	a medicine, or a substance for making medicines
6		antibiotic	N	a drug that is used to kill bacteria and cure infections.

Or a connected passage focused on target words

(‘far transfer’)

- For meaning
- For morphology
- For multiword units

VIRUSES AND BACTERIA

This complicated question, which should be answered by your healthcare provider, depends on the specific diagnosis. For example, there are seve of _____ - most need antibiotics, but some do not. Most cas sore throat are caused by viruses. One kind, strep throat, diagnosed by a _____, needs antibiotics.

Common _____, like coughs or a cold, can sometimes beco complicated and a _____ can develop. However, treating vir infections with antibiotics in order to prevent bacterial infections is not recommended because of the risk of causing _____:

bacterial infection
bacterial resistance
ear infections
lab test
viral infections

Remember that antibiotics do not work agains _____ e flu, an unnecessary antibiotics can be harmful.
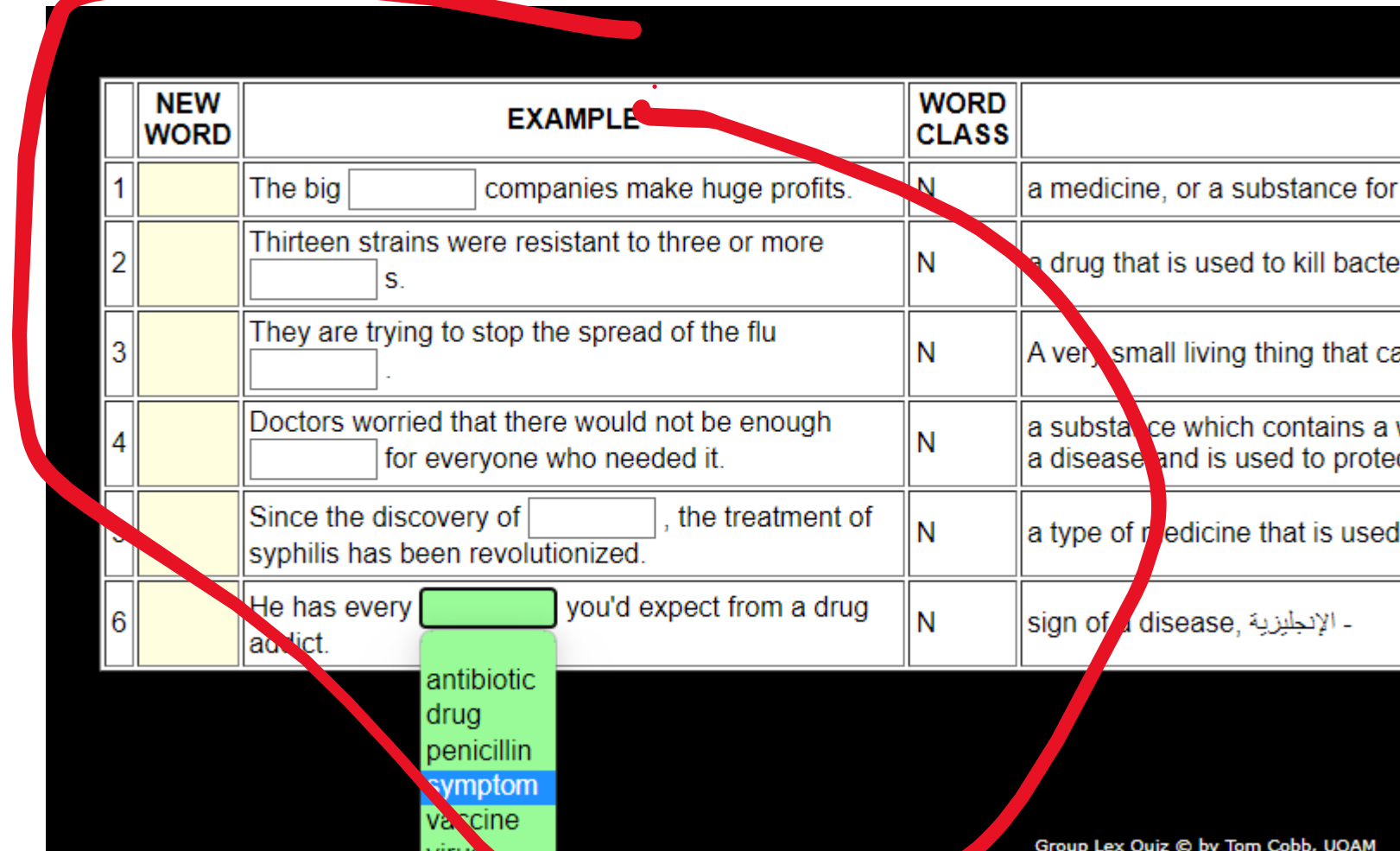
Talk with your healthcare provider about antib _____ about th differences between viruses and bacteria, and when antibiotics should a should not be used.

Gp Lex has elements of 'social media'

- Users show unexpected interest in doing quizzes with their classmates' words & examples (Oman and elsewhere)
  - Discussions overheard ~
    - "That is not the correct definition for this use of the word…"
    - "Your example does not make the word's meaning clear…"
    - "This is a verb not a noun"

But where do the 'comprehensible examples' for Group Lex come from?

	NEW WORD	EXAMPLE	WORD CLASS	
1		The big ☐ companies make huge profits.	N	a medicine, or a substance for r
2		Thirteen strains were resistant to three or more ☐ s.	N	a drug that is used to kill bacter
3		They are trying to stop the spread of the flu ☐ .	N	A very small living thing that cau
4		Doctors worried that there would not be enough ☐ for everyone who needed it.	N	a substance which contains a w a disease and is used to protec
5		Since the discovery of ☐ , the treatment of syphilis has been revolutionized.	N	a type of medicine that is used t
6		He has every ☐ you'd expect from a drug addict.	N	sign of a disease, الإنجليزية -

antibiotic
drug
penicillin
symptom
vaccine
viru

82

Larger context for **SYMPTOM** **Get**
in Corpus house/season_2.txt **more**

I knew I was forgetting something. You should do one of those while you're running the labs and the tox screen. The labs show Valium and heroin in her urine. A supermodel on smack. Shocker. Oh, Alex, I expected so much more from you. Heroin chic is so five years ago. **Okay, let's start crossing out withdrawal SYMPTOMs.** A positive test means she tried it once. It doesn't mean she's an addict. She's only 15. There's no age limit on addiction. He's right. She's never menstruated. **Sounds like a SYMPTOM of drug addiction to me.** Or bulimia. Or her age. Some girls don't start till their mid-to-late teens. Evidence to the contrary, the rounded hips. The perfectly-sculpted, bountiful breasts. Implants. I've seen some of her photos. They've grown dramatically since last summer. **SYMPTOMatic of turning 14.** Two clinic hours says that those "love apples" are

put (« Back keeps original settings) Copiable extract-link to this data >

ctionary/TTS Eng_Eng ∨

SORT ∨ VP ∨ ASSOC ON L|R ∨ FOR 25,000 ∨ L

rent SYMPTOM, a different diagnosis. Differ n [1.00]
's a SYMPTOM. It means the condition predat s [1.00]
ous SYMPTOM, you know? We'll know more afte [1.00]
's a SYMPTOM? Okay. So, what could cause ob s [1.00]
cond SYMPTOM. Yeah, yeah, kid's in V-tach. e [1.00]
arly SYMPTOM of MS. No. MR1 showed no white m [1.00]
's a SYMPTOM? Well, why would you think The b [1.00]
me a SYMPTOM that's not killing her. What el s [1.00]
the SYMPTOM. I'm wrong all the lime. Taking [1.00]
urth SYMPTOM, "being completely crazy." One [1.00]

So ~
We could go on, but perhaps the point is made

"DDL and the Arabic speaking learner is a particularly good match"

- Culture schemas are much reduced
- Meta-language is much reduced
- Vocab problem is met head on
- Contextual skills of L1 are exploited in concordance format
- Word perception issue from L1 is tackled (RT work)
- Speech-writing matches are assured (TTS work)
- Collaboration is built into the program
- Several choices are left to learners
- Transfer is targeted, not hit and miss

- Computers, phones, and social media are known motivators
- Guess-work in testing is eliminated or reduced
- Some drudgery is removed from teaching (quizzes, writing feedback) freeing up teachers for interaction
- All tests can be practiced in advance & to the extent desired
- Lextutor & related software is free & well supported (AntConc, Antprofiler)
- Most of these ideas have research support

References (1)

- Abu Rabia, S. & Siegel, L. (1995). Different orthographies different context effects: The effects of Arabic sentence context in skilled and poor readers. Reading Psychology 16 (1), 1-19.
- Arden-Close, C. (1993). Language problems in science lectures to non-native speakers. English for Specific Purposes 12, 251-261.
- Boulton, A. & Cobb, T. (2017). Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning 65 (2), 1-46.
- Cobb, T.M. (1995). Imported tests: Analyzing the task. Paper presented at TESOL (Arabia). Al-Ain, United Arab Emirates, March.
- Cobb, T. (2006). The old vocabulary, the new vocabulary, and the Arab learner. Text of plenary talk for Vocabulary Symposium, TESOL Arabia, Dubai, April 2006.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? System 25 (3), 301-315.
- Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. Educational Technology Research & Development, 47 (3), 15-33.

References (2)

- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? System, 32(3), 301-319.

- Meara & Buxton (1987). An alternative to multiple choice vocabulary tests. Language Testing, 4, 142-154)

- Nation, P. (2006). How many words do you need to be able to read? Canadian Modern Language Review.

- Nation, P. & Wang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? System 23(1), 35-41.

- Randall, M. & Meara, P. (1988). How Arabs Read Roman Letters. Reading in a Foreign Language, 4, 133-145.

- Ryan, A. & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. Reading in a Foreign Language 7(2), 531-540.