Data-driven learning - a little more colour

Evan Frendo

www.e4b.de

Abstract

Data-driven learning (DDL) has been around since the mid-eighties, and provides an excellent way for BE teachers to bring authentic language into the classroom. Yet despite its long history, some teachers remain hesitant to use it. In this session I will explain what DDL is and demonstrate just how useful it can be. The talk will include descriptions of easily available tools as well as activities which you can easily adapt to suit your own teaching context.

What is DDL (Data-Driven Learning)

Data-driven learning (DDL) consists in using the tools and techniques of corpus linguistics for pedagogical purposes.
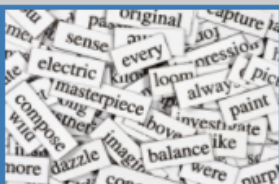
Gilquin and Granger, 2010, p. 359

Why DDL?

- Provide access to authentic language

- Check usage

- Encourage learners to explore language

- Multiple examples in context

- Develop learning skills e.g. noticing, reflecting, analyzing …

What is needed

- A corpus containing naturally occurring language data

- Tools to access the corpus (corpus analysis tools)

English-Corpora.org

The most widely used online corpora. **Tour (new: Nov 2020)**, overview, search types, variation, Virtual Corpora, corpus-based resources.

The links below are for the online interface. But you can also ⊙ download the corpora for use on your own computer.

Corpus (online access)	Download	# words	Dialect	Time period	Genre(s)
iWeb: The Intelligent Web-based Corpus	⊙	14 **billion**	6 countries	2017	Web
News on the Web (NOW)	⊙	11.3 **billion+**	20 countries	2010-yesterday	Web: News
Global Web-Based English (GloWbE)	⊙	1.9 **billion**	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	⊙	1.9 **billion**	(Various)	2014	Wikipedia
Corpus of Contemporary American English (COCA)	⊙	1.0 **billion**	American	1990-2019	Balanced
Coronavirus Corpus	⊙	694 million+	20 countries	Jan 2020-yesterday	Web: News
Corpus of Historical American English (COHA)	⊙	400 million	American	1810-2009	Balanced
The TV Corpus	⊙	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	⊙	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	⊙	100 million	American	2001-2012	TV shows
Hansard Corpus		1.6 **billion**	British	1803-2005	Parliament
Early English Books Online		755 million	British	1470s-1690s	(Various)
Corpus of US Supreme Court Opinions		130 million	American	1790s-present	Legal opinions
TIME Magazine Corpus		100 million	American	1923-2006	Magazine
British National Corpus (BNC) *		100 million	British	1980s-1993	Balanced
Strathy Corpus (Canada)		50 million	Canadian	1970s-2000s	Balanced
CORE Corpus		50 million	6 countries	2014	Web
From Google Books n-grams (compare)					
American English		155 billion	American	1500s-2000s	(Various)

THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學

Department of
ENGLISH
英文系 | Excelling in Professional Communication

Opening Minds • Shaping the Future
啟迪思維・成就未來

The PolyU Department of English
Language Bank

🏠 **Home** | 🔳 **Corpus Search** | ✉ **Contact Us**

Corpus Search

Category	Business ⌄
Corpus Data	Corpus of Business Correspondence ⌄
Part of Speech Tagging	**Corpus of Business Correspondence**
	Hong Kong Business Reports
	UK Business Reports
	US Business Reports
	Financial Crisis Corpus
	PolyU Business Corpus (English)
Search String	
Sort Type	

but, either) ⌄

SEARCH

Corpus File : PolyU Business Corpus (English)

Frequency : 591

```
1          y and improve consumer protection. Have " insurance '  money  on hand to cover changes Have " insurance ' money on
2           correction after eight days on the upside. "Overseas  money  is coming in . . . but I think it's time for the mar
3       n sufficient detail to show particulars of--- (a)      all  money  received or paid by the financier, including money p
4         , not from institutional hot money.'' ``Institutional  money  is not going to buy such tech issue,'' he said. ``In
5           elling Hong Kong stocks and going underweight. ``Less  money  left over to buy equity,'' he said, pointing out tha
6           nies) licensed under the MLO. The practice of using a  money  lender company to operate share margin financing bus
7           egulated save for the requirements that they obtain a  money  lenders licence under the MLO and thereafter comply
8            more than 5 per cent or 10 per cent? Frank Gannon, a  money  manager at SunAmerica Asset Management, sees a posit
9           erson is approved by the Monetary Authority (MA) as a  money  broker under the Ordinance. The Banking Ordinance pr
10          king Ordinance provides that no person shall act as a  money  broker unless that person is approved by the Monetar
11          of licensing and providing that any person, whether a  money  lender or not, who charges an interest rate in exces
12           a real honest dealer is that the scammer talks about  money  and the profit that you can make. To the real scamme
13          nds and having as its investment objective to achieve  money  market returns. The HSBC Fund invests in having as i
14           Money Market Fund is to preserve capital and achieve  money  The investment objective of the Money Market Fund is
15          research at Asia Financial Securities. ''It also adds  money  to PCCW's till.'' C&W HKT surged on news that its po
16          s as evidence of the advance. Where Mr. Wong advanced  money  without knowing the precise application of those fun
17      d by the Council.      (6)     (a)     Except as legally allowed,  money  held in a trust account shall not be available for p
18          the Money Market Fund directly in debt securities and  money  market investments rather than through an underlying
19          trading, clearing, settlement, share registration and  money  settlement functions, thereby reducing risk and the
20         the LER system. (II)   Conduct of foreign exchange and  money  market operations by the HKMA      (II)   Conduct of forei
21          ding, clearing and settlement, share registration and  money  settlement functions, with trade orders input by inv
22          ement facilities to the Australian fixed interest and  money  markets. ASX will improve its electronic trading sys
23          the first half of 1997, with the foreign exchange and  money  markets reacting calmly to the death of Mr. Deng Xia
24           have designed a wide range of flexible and value-for-money  products offering protection and savings through who
25          KMA's occasional presence in the foreign exchange and  money  markets does not seem to be clearly understood. Cert
```

The EnronSent Corpus

The EnronSent corpus is a special preparation of a portion of the Enron Email Dataset designed specifically for use in Corpus Linguistics and language analysis. Divided across 45 plain text files, this corpus contains 2,205,910 lines and 13,810,266 words.

AntConc 3.5.8 (Windows) 2019 — □ ✕

File   Global Settings   Tool Preferences   Help

Corpus Files

enronsent00.txt
enronsent01.txt
enronsent02.txt
enronsent03.txt
enronsent04.txt
enronsent05.txt
enronsent06.txt
enronsent07.txt
enronsent08.txt
enronsent09.txt
enronsent10.txt
enronsent11.txt
enronsent12.txt
enronsent13.txt
enronsent14.txt
enronsent15.txt
enronsent16.txt
enronsent17.txt
enronsent18.txt
enronsent19.txt
enronsent20.txt
enronsent21.txt
enronsent22.txt
enronsent23.txt
enronsent24.txt
enronsent25.txt
enronsent26.txt
enronsent27.txt
enronsent28.txt
enronsent29.txt

Total No.

45

Files Processed

| Concordance | Concordance Plot | File View | Clusters/N-Grams | Collocates | Word List | Keyword List |

Concordance Hits   1432

Hit	KWIC	File
675	a score of 68. The mid-year RED index also found that 62% of states	enronsent20
676	about obtaining a copy of the RED Index, go to www.caem.org.	enronsent20
677	utilities to go $12 billion into the red. California's governor Gray Davis responded	enronsent20
678	release with our initial changes in red. Please review and let me know	enronsent20
679	Ariba lost 5.13 to 152.44, both on Nasdaq. Red Hat (Nasdaq) lost 4.06 to 21.19. The Research	enronsent20
680	give you an idea how the red/green color coding for rating movement	enronsent21
681	to the Retail Energy Deregulation Index (RED Index), published twice a year by	enronsent21
682	due to increasing wholesale prices. The RED Index, which is the only known	enronsent21
683	a score of 68. The mid-year RED Index also found that 62% of states	enronsent21
684	about obtaining a copy of the RED Index, go to www.caem.org.	enronsent21
685	, another Midwest service center exec counte= red, has got nothing to do with	enronsent21
686	quickly picked up my copy of Red Herring yesterday to read the Enron	enronsent21
687	Somebody PLEASE help me. How can Red Herring POSSIBLY give advice to Enron?!!!!	enronsent21
688	Herring POSSIBLY give advice to Enron?!!!! Red Herring's advertisement revenues in July	enronsent21
689	it Louise, Your approval on the red color will help us proceed quickly.	enronsent21
690	you can see how prominent the red one is in our home. Take	enronsent21
691	: NETCO EEI COLLATERAL ANNEX COVER SHEET RED DOC >> << File: NETCO EEI Collateral Annex	enronsent21

Search Term ☑ Words ☐ Case ☐ Regex

red   [Advanced]

Search Window Size

50

[Start] [Stop] [Sort] **Show Every Nth Row** 1

Kwic Sort

☐ Level 1 1R ☐ Level 2 2R ☐ Level 3 3R   [Clone Results]

Red

A copy of the agreement is kept in a red file (near Sam's desk)

Attached is the red-line version of the …

So it's a red-letter day for us …

I will raise a red flag …

He might be barking at a red herring.

My answers below in red.

Take the red-eye flight home on Monday night.

So let me know if I have any money coming to me or if I am in the red.

… investors fear more red ink splashed over the balance sheets

I drank more red wine than beer that weekend.

The paperwork and red tape have seemed unending …

I've already marked up w/ red pen …

Specialist language – *aground*

. What part is aground? Aground forward. Aground amidships. Aground aft. Aground full length.
is aground ? Vessel aground forward. Vessel aground amidships. Vessel aground aft. Vessel aground
in collisions or contacts, around 18% ran aground and around 12% had fires on board.
attacked, she could always be run aground, and her more than 10,000 passengers could
AGROUND - A vessel which has gone aground and is incapable of re-floating
36000 gt bulk carrier Fedra, which ran aground and lodged under a cliff after
2003, the general cargo vessel Jambo ran aground and subsequently sank off the west
on water. Realising that Surprise was aground and taking on water, the skipper
that the ship would shortly be aground and that the starboard swing needed
and Explosion   If the ship is aground and therefore cannot manoeuvre, all possible
passenger ship Giorgis, which ran hard aground and was holed off the island
vehicles on board shifted, it ran aground and was left high and dry
ability to manoeuvre or is firmly aground. As the salvage costs are covered
aground at high water. I went aground at half water. I went aground
in position ... require assistance. I went aground at high water. I went aground
"Aegean Sea"),60 the Aegean Sea went aground at La Coruña in Spain on 3

DDL and learner autonomy
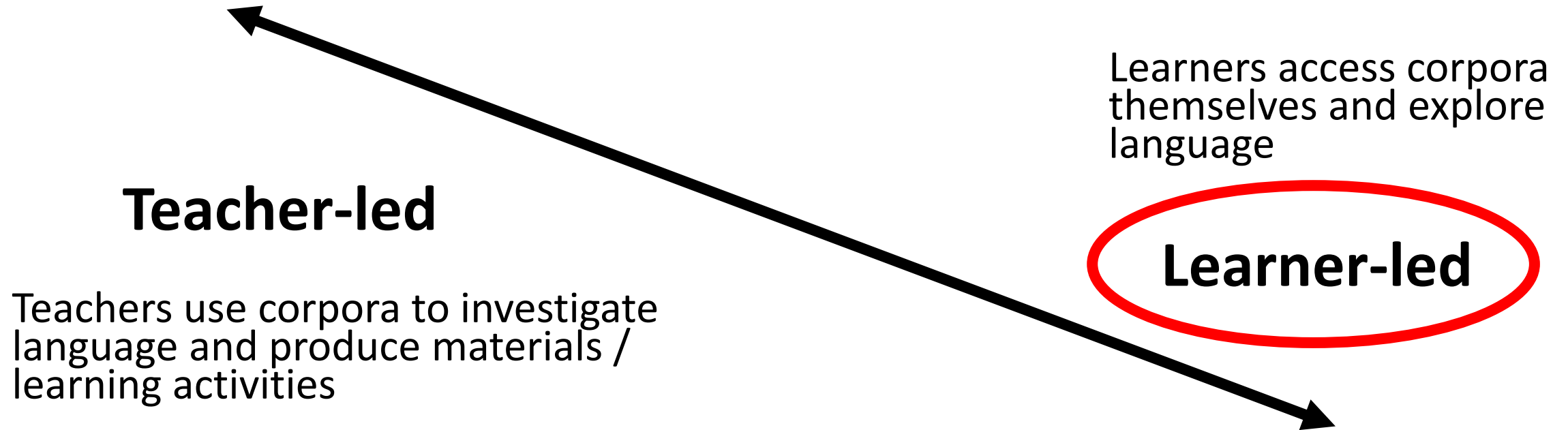
Teacher-led

Teachers use corpora to investigate language and produce materials / learning activities

Learners access corpora themselves and explore language

Learner-led

Mukherjee, 2006

SkELL

Sketch Engine for Language Learning

Discover the English language through a billion-word collection of news, scientific papers, Wikipedia articles, fiction, web pages and blogs.

Examples: search for a word or a phrase and get short and meaningful example sentences for it.

Word sketch is a list of words which occur frequently together with the searched word.

Similar words (not only synonyms) are words used in similar contexts visualized with a word cloud.

Sketch Engine is a state-of-the-art cloud tool for building, managing and exploring large text collections in dozens of languages. It is used all over the world by many individuals, as well as companies such as Cambridge University Press, Oxford University Press and Macmillan.

SkELL

information   🔍

information   394.24 hits per million

1. **Information** without visceral context is neither useful nor retained.

2. All final disciplinary actions involving teacher licenses are public **information** .

3. This aggregate **information** includes anonymous website statistics.

4. Create **information** goods regarding present hot topics.

5. Your personally identifiable **information** is kept secure.

6. But getting more **information** is proving difficult.

7. This **information** remains valid until another event is handled.

8. Popular government without popular **information** is impossible.

9. This means faster more detailed **information** is available.

10. A little background **information** is helpful too.

SkELL

information — search box

information [noun]   Context ⬤

verbs with information as subject

| regard | concern | relate | please | pertain | gather | need | provide | contain | present | indicate | collect | obtain | include | gain |

verbs with information as object

| provide | collect | contain | gather | obtain | share | disclose | store | receive | request | give | get | present | need | find |

adjectives with information

| available | relevant | confidential | online | necessary | accessible | accurate | correct | useful | specific | secure | pertinent | such |

| private | vital |

modifiers of information

| more | personal | additional | further | detailed | contact | useful | relevant | much | confidential | valuable | background | such |

| accurate | sensitive |

Search...

- Use ***** to match zero or more words, e.g., `play * role` -> `play a role` and `play an important role`

- Use **_** to match any word, e.g., `listen _ music` -> `listen to music`

- Use **~WORD** to search for the similar words of a WORD, e.g., `~reliable person` -> `trustworthy person`

- Use **?WORD** to search for WORD optionally, e.g., `listen ?to music` -> `listen to music`

- Use **WORD1/WORD2 (slash)** to choose either WORD1 or WORD2 e.g., `go in/to school` -> `go to school`

- Use **POS (Part of Speech)** to search for word with specific pos tag, e.g., `v. a report` -> `submit a report`
  POS can be: `v. (verb)` `n. (noun)` `adj. (adjective)` `prep. (preposition)` `det. (determiner)` `conj. (conjunction)` `pron. (pronoun)`
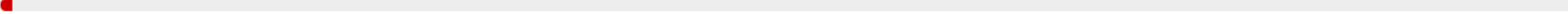 Remember to keep a dot(.) after each pos tag

Linggle.com

N-gram	Percent	Count	Example
⚠ much money	99.3 %	1,800,000	+
many money	0.7 %	13,000	+

N-gram	Percent	Count	Example
⚠ much money	99.3 %	1,800,000	—

- How Much Money?/ by Hollie J. [Google Books]
- Remember, Lil gave all that eighteenth-century French furniture to the Metropolitan Museum, and they named the room after her, back when she still had so much money. [Google Books]
- The big question in our lives is not, "How much money is enough? [Google Books]
- How Much Money? [Google Books]
- Distribute copies of How Much Money? to students. [Google Books]
- It includes a general section on getting to know PowerPoint: sections on creating and customizing your first presentation Find Out How Much Money You Will Need to Run. [Google Books]
- How much money does Lenny have left? [Google Books]
- Too Much Money . [Google Books]
- How much money? [Google Books]
- P2D3 Total: _____ cents How much money? [Google Books]
- I worried about how much money we were spending . [COCA]

N-gram	Percent	Count	Example
⚠ much money	99.3 %	1,800,000	+
many money	0.7 %	13,000	−

- ... you spend I time How many money repair can ... [Google Books]
- Many money gurus see this differently, but it's a time in history like no other. [Google Books]
- (1996) have tabulated some of many money questionnaires developed by researchers and the possible factors that influenced them (see Table 2.4) . [Google Books]
- Many money experts agree that at every income level, people fall into one of the following three categories: ✓ People who spend more than they earn (accumulating debt) ✓ People who spend all that they earn (saving nothing) ✓ People who ... [Google Books]
- Choose to Save provides articles and calculators for thinking about many money-related situations. [Google Books]
- Uniting programs to focus on the same objective would thwart many money launderers attempting to navigate around the system ... [Google Books]
- Finally, I would like to thank the many Money and Banking instructors who contributed their expertise and ideas by reviewing the proposal and drafts of the manuscript, attending focus groups, replying to surveys, and participating in Web ... [Google Books]
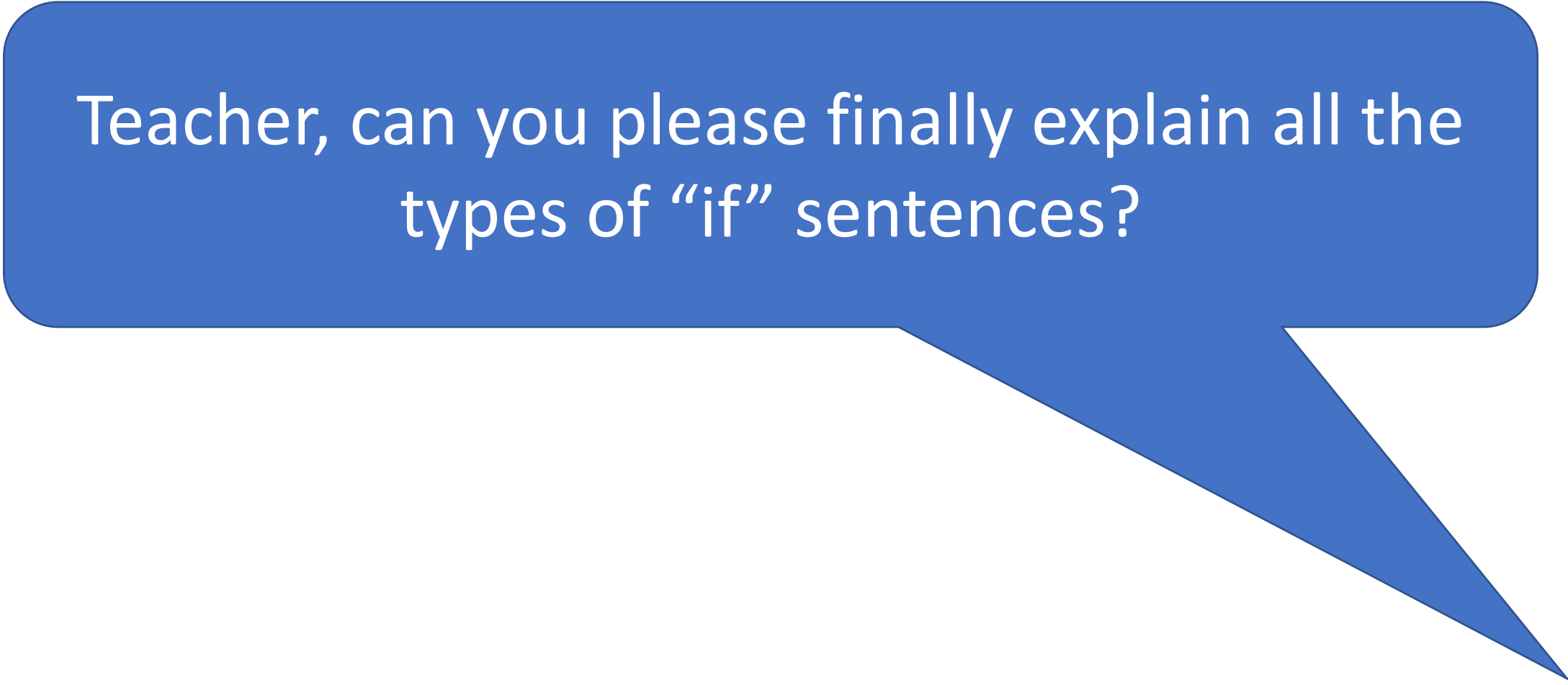- But itis not surprising thatthe Americans haveproduced so many money innovators.

Netspeak

One word leads to another.

i ✕ ↺

how to ? this	The ? finds one word.
see ... works	The ... finds many words.
it's [great well]	The [ ] compare options.
and knows #much	The # finds similar words.
{ more show me }	The { } check the order.
m...d ? g?p	The space is important.

Teaching conditionals

- Zero (If + present simple, present simple),

- First (If + present simple, will + infinitive),

- Second (If + past simple, would + infinitive), and

- Third (If + past perfect, would have + past participle)

- Mixed conditionals

If you pop over there: a corpus-based study of conditionals in medical discourse

Gibson Ferguson*

University of Edinburgh, Institute for Applied Language Studies, 21 Hill Place, Edinburgh EH8 9DP, UK

Conditionals revisited

Glenn Fulcher

Using a reasonably large database of written text and a simple statistical technique, this small study looks once again at the issue of teaching conditionals and other if forms. It is suggested that the selection of the forms to be included in the syllabus may be related to the needs of the students in studying English, while at the same time acknowledging that certain aspects of the results are restricted in application and may benefit from a study using a much more precise definition of text-type.

'Sorry, but if he comes, I go': teaching conditionals

David Maule

This article looks at the rather restricted selection of conditional sentences which tends to be presented to students. It cites one case of the damage to a student's natural development which this might cause and, on the basis of a fairly sm... of these... which i... gramma...

If only it were true: the problem with the four conditionals

Christian Jones and Daniel Waller

The traditional division of conditionals into four main types (zero, first, second, and third) has long been called into question. Unfortunately, the awareness that this description does not reflect conditional patterns in actual usage has not generally been reflected in EFL coursebooks. This article re-examines the arguments for a description of conditional patterns which reflects actual usage and uses corpus data to demonstrate the kind of patterns in frequent use. It then suggests two teaching approaches that may help teachers to tackle a variety of conditional patterns in the classroom.
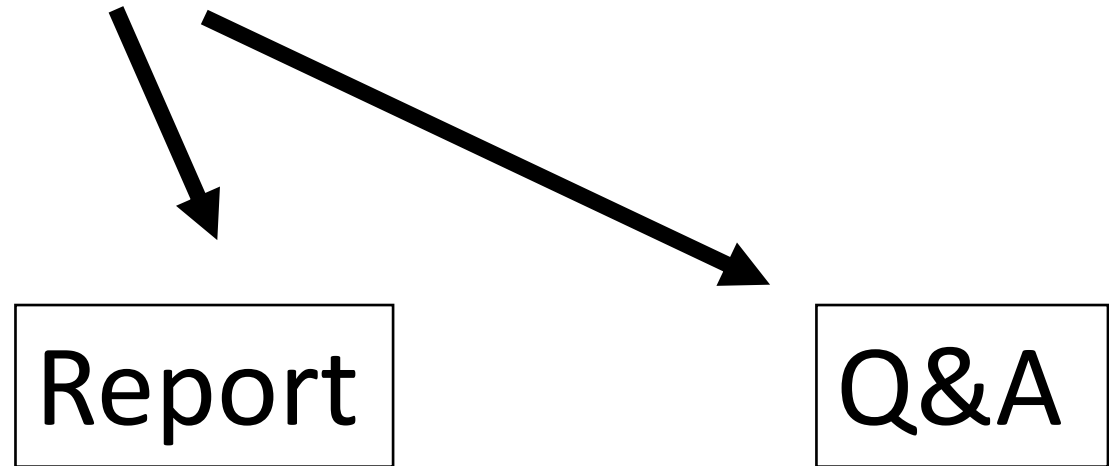
Earnings calls transcripts

Conference calls during which company executives present periodic financial results to investment analysts connected via telephone and the Internet

Crawford Camiciottoli, 2010, p 343

Earnings calls transcripts

Report   Q&A

- Easily available online
- Often include recordings as well as transcripts

1. And if so, what's the major roadblock for others getting involved?

2. And sorry, just if I can squeeze one more.

3. I'm not sure if you put out there what the number is today.

4. If you look at sort of what your burn was, the last couple of quarters, would that kind of put where your balance is today?

5. I'm just wondering if that's appropriate, if you can comment on that.

6. If this product is better, your average pricing should continue to trend down.

7. I'm wondering if you could provide any guidance with respect to your product mix.

8. So, if you could just give any color with respect to what your position on that is.

9. So, if you're wanting to pay less you get rid of all those mistakes.

10. I just have one follow-up if you don't mind?

11. I don't know if it will stay that way.

12. If you use our technology, you don't get a hundred new features, every quarter.

13. I just wonder if you could comment on just sort of the geographic landscape for you all.

14. I was wondering if you just add a little bit color on sort of what you're seeing across the different markets.

15. If you did not receive a copy, you may obtain the information from the website.

16. Can you talk about if there is any impact there of any investments, you might be considering?

17. I don't know if that's what you were trying to reference or allude to

18. If some of the tariffs are rolled back, how is the company going to handle this?

19. I don't know if we should be thinking about something similar on a go-forward basis?

1. And if so, what's the major roadblock for others getting involved?

2. And sorry, just if I can squeeze one more.

3. I'm not sure if you put out there what the number is today,

4. I...

5. I'... comment on that.

6. If this product is better, your average pricing should continue to trend down.

7. I'm wondering if you could provide any guidance with respect to your product mix,

8. So, ...ce or you...

9. So, if you're wanting to pay less you get rid of all those mistakes.

10. **I just have one follow-up if you don't mind?**

11. I don't know if it will stay that way.

12. **If you use our technology, you don't get a hundred new features, every quarter.**

13. I just wonder if you could comment on just sort of the ... information from the website.

16. Can you talk about if there is any impact there of any investments, you might be considering?

...any going to handle this?

19. I don't know if we should be thinking about something similar on a go-forward basis?

1. And if so, what's the major roadblock for others getting involved?

2. And sorry, just if I can squeeze one more.

3. I'm not sure if you put out there what the number is today.

4. If you look at sort of what your burn was, the last couple

5.

6.

7. I'm wondering if you could provide any guidance with respect to your product mix.

8. So, if you could just give any color with respect to what your position on that is.

9. So, if you're wanting to pay less you get rid of all those mistakes.

10. I just have one follow-up if you don't mind?

11. **I don't know if it will stay that way.**

12. If you use our technology, you don't get a hundred new features, every quarter.

13. I just wonder if you could comment on just sort of the geographic landscape for you all.

investments, you might be considering?

17. I don't know if that's what you were trying to reference or allude to

18. If some of the tariffs are rolled back, how is the company going to handle this?

19. I don't know if we should be thinking about something similar on a go-forward basis?

11. I don't know if it will stay that way.

1. And if so, what's the major roadblock for others getting involved?

2. And sorry, just if I can squeeze one more.

3.

4.

7. ==I'm wondering if you could provide any guidance with respect to your product mix.==

5. I'm just wondering if that's appropriate, if you can comment on that.

6. If this product is better, your average pricing should continue to trend down.

7. ==I'm wondering if you could provide any guidance with respect to your product mix.==

8. So, if you could just give any color with respect to what your position on that is.

9. So, if you're wanting to pay less you get rid of all those mistakes.

10. I just have one follow-up if you don't mind?

11. I don't know if it will stay that way.

12. If you use our technology, you don't get a hundred new features, every quarter.

15. If you did not receive a copy, you may obtain the information from the website.

16. Can you talk about if there is any impact there of any investments, you might be considering?

17. I don't know if that's what you were trying to reference or allude to

18. If some of the tariffs are rolled back, how is the company going to handle this?

19. I don't know if we should be thinking about something similar on a go-forward basis?

1. And if so, what's the major roadblock for others getting involved?

2. And sorry, just if I can squeeze one more.

3. I'm not sure if you put out there what the number is today.

8. **So, if you could just give any color with respect to what your position on that is.**

6. If this product is better, your average pricing should continue to trend down.

7. I'm wondering if you could provide any guidance with respect to your product mix.

8. So, if you could just give any color with respect to what your position on that is.

9. So, if you're wanting to pay less you get rid of all those mistakes.

10. I just have one follow-up if you don't mind?

11. I don't know if it will stay that way.

12. If you use our technology, you don't get a hundred new features, every quarter.

13. I just wonder if you could comment on just sort of the geographic landscape for you all

information from the website.

16. Can you talk about if there is any impact there of any investments, you might be considering?

17. I don't know if that's what you were trying to reference or allude to

18. If some of the tariffs are rolled back, how is the company going to handle this?

19. I don't know if we should be thinking about something similar on a go-forward basis?

color - requests for additional information

- Any color there would be helpful.
- To add a little more color, …
- I have no real color to add.
- I guess I'd add just a little bit of color.
- You've given some good color on how you think the markets will …
- If you could provide some color on …
- Could you talk about what sort of color …
- Appreciate the color.
- Thanks for the color.

A final thought

What is distinctive about the DDL approach to inductive language teaching is the principle that the data is primary, and the teacher does not know in advance exactly what rules or patterns the learners will discover: indeed, they will often notice things that are unknown not only to the teacher, but also to the standard works of reference on the language. It is this element of challenge and of discovery that gives DDL its special flavour and stimulus.

Johns, 1991, p. 2

References

Crawford Camiciottoli, B. (2010). Earnings calls: Exploring an emerging financial reporting genre. *Discourse and Communication, 4*(4), 343-359.

Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse, *English for Specific Purposes,* 20, 61–82.

Frendo, E. (2020). 'If you could just comment on what you're seeing?' Some thoughts on the teaching of conditionals and the use of data-driven learning with business English learners. *ETAS Journal, 37* (2), 31-32.

Fulcher, G. (1991). Conditionals revisited. *ELT Journal*, *45*(2), 164–68.

Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*, London, England: Routledge, (pp. 359–370).

Johns, T. (1991). 'Should You Be Persuaded: Two examples of data-driven learning'. In T. Johns and P. King (Eds.), *Classroom concordancing, ELR Journal 4*. Birmingham, England: Centre for English Language Studies, University of Birmingham, (pp. 1–16).

Jones, C. and Waller, D. (2011). If only it were true: the problem with the four conditionals. *ELT Journal*, *65*(1), 24–32.

Maule, D. (1988). 'If he comes, I go': teaching conditionals. *ELT Journal*, *42*(2), 117–23.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In S. Braun, K. Kohn and J. Mukherjee (Eds.), *Corpus technology and language pedagogy*. Frankfurt am Main, Germany: Peter Lang, (pp. 5–24).

Corpora / corpus analysis tools

AntConc https://www.laurenceanthony.net/software/antconc/

English Corpora https://www.english-corpora.org/

Enron Email dataset https://wstyler.ucsd.edu/enronsent/

Linggle https://linggle.com/

Netspeak https://netspeak.org/

Research Centre for Professional Communication In English http://rcpce.engl.polyu.edu.hk/rcpce/corporaresources.html

SKELL https://www.sketchengine.eu/skell/