Corpus Linguistics
basic concepts and methods
Essentials

- How do we study language?
- Why do we study language?
Essentials

- How do we study language?

- Why do we study language?
  - ‘Applied’ linguistics
Studying language

- ‘think’ about language and make up examples
- listen to others
- elicit data
- look at any old text?
- The professor’s shoeboxes

- With corpus linguistic methods: what kind of ‘texts’ and how do we look at them?
Corpus Linguistics

- investigates language on the basis of electronically stored samples of naturally occurring language

- ‘corpus’ is a collection of such language samples stored in a principled way in order to address linguistic questions
Corpora

- Depending on the research question, corpora contain: newspapers, leaflets, transcriptions of spoken language, political speeches, text messages, fiction, videos, sign language, etc.

- Principled or opportunistic? (cf. Web as corpus)
- Can be processed in a variety of ways
- Language awareness: most frequent noun in English?
Try  http://www.wordle.net
Possible research questions

- Do men or women swear more?
Possible research questions

- Do men or women swear more?
- When did *nerd* enter the language?
Possible research questions

- Do men or women swear more?
- When did *nerd* enter the language?
- How have the connotations of *nerd* and *geek* changed over time?
Corpus and Discourse

- Language is a social phenomenon
- Discourse as representation: words label participants and entities, express world views, etc.
Diane Abbott to warn of British 'masculinity crisis'

There is a "crisis of masculinity in Britain" because of the pressures rapid economic and social change have placed on masculine identity, shadow health minister Diane Abbott is to claim.

In a speech on Thursday, Ms Abbott will warn a generation of men are in transit and unclear of their social role.

They are also under pressure to live up to "pornified ideals", with Viagra becoming a party drug, she will say.

She will urge more action to help boys see a less narrow view of masculinity.

Speaking at the think tank Demos in London, Ms Abbott will set out how she thinks changing social attitudes and movements in the labour market have left men "isolated and misdirected".

But she will stress that the problems men and boys face are not receiving sufficient attention because - "like the film Fight Club - the first rule of being a man in modern Britain is that you're not allowed to talk about it".

'Hyper-masculinity'
Corpus and Discourse

- Corpus design and compilation (e.g. *sustainable development* Corpus)
- ‘seed words’
Display formats

- Word clouds
- Tables
- Figures
Display formats

- Word clouds / Frequency lists
- Tables
- Figures
- Concordances

“The language looks rather different when you look at a lot at once.”

(John Sinclair 1991)
News Motorist fined for blowing nose vigorously exercise at 7.9% and nose blowing becomes internet hit A motorist was fined by vigorous exercise at 7.9% and nose blowing at 5.4%. These were
News Video Baby scared of nose blowing becomes internet hit A Andrew tries to keep his nose clean in Indonesia By Andy
Pushing envelopes Pushing envelopes: red nose has to pay through the nose day Pride and embarrassment for using tissue at the Welsh are paying through the nose for insurance' Do you live for Sir William Castell at home Peter Simon shows his nose for a good bargain on free Andy Coulson had a nose for the view of the Things Are: The shoebill - a nose for the unusual By Mike Japan Elderly woman breaks student's nose for failing to give up Doctor, doctor: Why does my nose get blocked? 'I worry this Snooker: No skin off snooker's nose if Rocket quits, says Davis Surgery: Ed votes for a nose job Miliband hopes for better Dianna Agron admits to having nose continued with plastic surgery' after nose job patient dies from heart job? He's set to wed Does Mike Tindall need a nose volume reveals brutality of early nose job 19th century book for that he has a red nose: like all of us, he's
Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World Ed Miliband undergoes nose operation for sleep apnoea Politics Ed Miliband Ed Miliband nose operation satisfies surgeons Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July Ed Nose Day' Ed Miliband's nose operation could prove vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked job patient dies from heart job? He's set to wed Does Mike Tindall need a nose volume reveals brutality of early nose job 19th century book for that he has a red nose: like all of us, he's
Science News Having a big nose prevents sniffles Having a Great Dane bites off courier's nose The owner of two barking 1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose A motorist was fined by vigorous exercise at 7.9% and nose blowing at 5.4%. These were
News Video Baby scared of nose blowing becomes internet hit A
Andrew tries to keep his nose clean in Indonesia By Andy
Pushing envelopes Pushing envelopes: red nose has to pay through the nose for using tissue at the
Welsh are paying through the nose for insurance' Do you live for a good bargain on
Oil and Gas A bloody nose Peter Simon shows his nose for the view of the
home Peter Simon shows his nose free Andy Coulson had a nose
Things Are: The shoebill - a nose for the unusual By Mike
Japan Elderly woman breaks student's nose for failing to give up
Doctor, doctor: Why does my nose get blocked? 'I worry this
Snooker: No skin off snooker's nose if Rocket quits, says Davis
Surgery: Ed votes for a nose job Miliband hopes for better
Dianna Agron admits to having nose job Glee star Dianna Agron continued with plastic surgery' after nose job patient dies from heart
Does Mike Tindall need a nose job? He's set to wed
volume reveals brutality of early nose that he has a red nose: like all of us, he's
Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World
Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea
Politics Ed Miliband Ed Miliband nose operation satisfies surgeons
Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband
UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July
Ed Nose Day' Ed Miliband's nose operation could prove vigilant
Phoenix Jones suffers broken nose Phoenix Jones, the masked
Science News Having a big nose prevents sniffles Having a
Sunday | Home More on: runny nose
Great Dane bites off courier's nose The owner of two barking
1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose vigorous exercise at 7.9% and nose blowing becomes internet hit A motorist was fined by clean in Indonesia By Andy Andrew tries to keep his nose day Pride and embarrassment

Pushing envelopes Pushing envelopes: red nose has to pay through the nose for using tissue at the Welsh are paying through the nose for insurance' Do you live

Oil and Gas A bloody nose for Sir William Castell at home Peter Simon shows his nose for a good bargain on free Andy Coulson had a nose for the view of the Things Are: The shoebill - a nose for the unusual By Mike Japan Elderly woman breaks student's nose for failing to give up

Doctor, doctor: Why does my nose get blocked? 'I worry this Snooker: No skin off snooker's nose if Rocket quits, says Davis Surgery: Ed votes for a nose job Miliband hopes for better job Glee star Dianna Agron continued with plastic surgery' after nose job patient dies from heart job? He's set to wed

Does Mike Tindall need a nose volume reveals brutality of early nose that he has a red nose: like all of us, he's jobs 19th century book for Advert屁ment sport Meagan Duhamel breaks nose of Eric Radford at World

Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea Politics Ed Miliband Ed Miliband nose operation satisfies surgeons Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July Ed Nose Day' Ed Miliband's nose operation could prove vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked Science News Having a big nose prevents sniffles Having a

Sunday | Home More on: runny nose Great Dane bites off courier's nose The owner of two barking 1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose vigorously at 7.9% and nose blowing at 5.4%. These were
News Video Baby scared of nose blowing becomes internet hit A clean in Indonesia By Andy
Andrew tries to keep his nose day Pride and embarrassment
Pushing envelopes Pushing envelopes: red nose has to pay through the nose
for using tissue at the
Welsh are paying through the nose for insurance' Do you live for a good bargain on
home Peter Simon shows his nose for the view of the
free Andy Coulson had a nose for the unusual By Mike
Things Are: The shoebill - a nose Japan Elderly woman breaks student's nose for failing to give up
Doctor, doctor: Why does my nose get blocked? 'I worry this
Snooker: No skin off snooker's nose if Rocket quits, says Davis
Surgery: Ed votes for a nose job Miliband hopes for better
Dianna Agron admits to having nose job patient dies from heart
continued with plastic surgery' after nose job? He's set to wed
Does Mike Tindall need a nose volume reveals brutality of early nose jobs 19th century book for
that he has a red nose: like all of us, he's
Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World
Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea
Politics Ed Miliband Ed Miliband nose operation satisfies surgeons
Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband
UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July
Ed Nose Day' Ed Miliband's nose operation could prove
vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked
Science News Having a big nose prevents sniffles Having a
Sunday | Home More on: runny nose
Great Dane bites off courier's nose The owner of two barking
1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose vigorously at 7.9% and nose blowing at 5.4%. These were News Video Baby scared of nose blowing becomes internet hit A clean in Indonesia By Andy Pushing envelopes: red nose has to pay through the nose day Pride and embarrassment for using tissue at the Welsh are paying through the nose for insurance' Do you live Oil and Gas A bloody nose for Sir William Castell at home Peter Simon shows his nose for a good bargain on free Andy Coulson had a nose for the view of the Things Are: The shoebill - a nose for the unusual By Mike Japan Elderly woman breaks student's nose for failing to give up Doctor, doctor: Why does my nose get blocked? 'I worry this Snooker: No skin off snooker's nose if Rocket quits, says Davis Surgery: Ed votes for a nose job Miliband hopes for better Dianna Agron admits to having nose job Glee star Dianna Agron continued with plastic surgery' after nose job patient dies from heart Does Mike Tindall need a nose job? He's set to wed volume reveals brutality of early nose jobs 19th century book for that he has a red nose: like all of us, he's Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea Politics Ed Miliband Ed Miliband nose operation satisfies surgeons Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July Ed Nose Day' Ed Miliband's nose operation could prove vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked Science News Having a big nose prevents sniffles Having a Sunday | Home More on: runny nose Refine: immune hay fever Great Dane bites off courier's nose The owner of two barking 1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose vigorous exercise at 7.9% and nose blowing at 5.4%. These were Andrew tries to keep his nose clean in Indonesia By Andy Pushing envelopes Pushing envelopes: red nose has to pay through the nose for using tissue at the Welsh are paying through the nose for insurance' Do you live Oil and Gas A bloody nose for Sir William Castell at home Peter Simon shows his nose for a good bargain on free Andy Coulson had a nose for the view of the Things Are: The shoebill - a nose for the unusual By Mike Japan Elderly woman breaks student's nose for failing to give up Doctor, doctor: Why does my nose get blocked? 'I worry this Snooker: No skin off snooker's nose if Rocket quits, says Davis Surgery: Ed votes for a nose job Miliband hopes for better Dianna Agron admits to having nose job Glee star Dianna Agron continued with plastic surgery' after nose job patient dies from heart Does Mike Tindall need a nose job? He's set to wed volume reveals brutality of early nose jobs 19th century book for that he has a red nose: like all of us, he's Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea Politics Ed Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July Ed Nose Day' Ed Miliband's nose operation could prove vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked Science News Having a big nose prevents sniffles Having a Sunday | Home More on: runny nose Refine: immune hay fever Great Dane bites off courier's nose The owner of two barking 1837: Woman bites off man's nose The Guardian reports on a
News Motorist fined for blowing nose vigorously at 7.9% and nose blowing at 5.4%. These were News Video Baby scared of nose blowing becomes internet hit.

Andrew tries to keep his nose clean in Indonesia By Andy

Pushing envelopes: red nose has to pay through the nose for using tissue at the day Pride and embarrassment

Welsh are paying through the nose for insurance' Do you live for a good bargain on

Oil and Gas A bloody nose for Sir William Castell at

home Peter Simon shows his nose free Andy Coulson had a nose for the view of the

Things Are: The shoebill – a nose for the unusual By Mike

Japan Elderly woman breaks student's nose for failing to give up

Doctor, doctor: Why does my nose get blocked? 'I worry this

Snooker: No skin off snooker's nose if Rocket quits, says Davis

Surgery: Ed votes for a nose job Miliband hopes for better

Dianna Agron admits to having nose job Glee star Dianna Agron

continued with plastic surgery' after nose job patient dies from heart

Does Mike Tindall need a nose job? He's set to wed

volume reveals brutality of early nose jobs 19th century book for

that he has a red nose: like all of us, he's

Advertisement sport Meagan Duhamel breaks nose of Eric Radford at World

Ed Miliband Ed Miliband undergoes nose operation for sleep apnoea

Politics Ed Miliband Ed Miliband nose operation satisfies surgeons

Miliband Ed Miliband to undergo nose operation on NHS Ed Miliband

UK Politics Ed Miliband undergoes nose operation Wednesday, 27 July

Ed Nose Day' Ed Miliband's nose operation could prove

vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked

Science News Having a big nose prevents sniffles Having a

Sunday | Home More on: runny nose Refine: immune hay fever

Great Dane bites off courier's nose The owner of two barking

1837: Woman bites off man's nose The Guardian reports on a
Concordances

- Show patterns
- Used to make dictionaries:
  
  blow your nose
  pay through the nose
  nose job

  Ed Miliband undergoes nose operation
Concordances

- Show patterns
- Used to make dictionaries:
  - blow your nose
  - pay through the nose
  - nose job
  - Ed Miliband undergoes nose operation
- Representativeness
WebCorp Live lets you access the Web as a corpus - a large collection of texts from which examples of real language use can be extracted. More...

Search: nose

Case Insensitive: ✓
Span: 50 characters

Search API: Google
Language: Not specified

Show URLs: ✓
Pages: 64

One concordance line per web page: 

Site: thetimes.co.uk telegraph.co.uk guardian.co.uk observer.co.uk independent.co.uk
WebCorp Live lets you access the Web as a corpus - a large collection of texts from which examples of real language use can be extracted. More...

Search:

Case Insensitive: ✓ Span: 50 characters

Search API: Google

Show URLs: ✓ Pages: 64

One concordance line per web page: 

Site:

thetimes.co.uk telegraph.co.uk guardian.co.uk observer.co.uk independent.co.uk
WebCorp Live lets you access the Web as a corpus - a large collection of texts from which examples of real language use can be extracted. More...

Search: nose

Case Insensitive: ✓

Span: 50 characters

Search API: Google

Language: Not specified

Show URLs: ✓

Pages: 64

One concordance line per web page: 

Site: thetimes.co.uk telegraph.co.uk guardian.co.uk observer.co.uk independent.co.uk
The empirical basis of corpus linguistics

- Emphasis on social dimension of language
- Electronic format that allows further procession with computer tools
Corpus linguistics in context

- focus on the text and the language, not the technology
- cf. NLP, computational linguistics
- wider context of Digital Humanities
Who uses corpora and corpus methods:
Who uses corpora and corpus methods:

- Lexicographers, teachers, …
- Political sciences: White House press briefings, news reporting on Iraq war
- Consultancies for companies: brands represented in the news
- Language of conflict managements
- Subtitling of films
Who uses corpora and corpus methods:

- Lexicographers, teachers, ...
- Political sciences: White House press briefings, news reporting on Iraq war
- Consultancies for companies: brands represented in the news
- Language of conflict managements
- Subtitling of films
s flagship Odeon cinema said: `I am afraid Ally's been subbed.' <p> Rangers notions in ten years. I'm afraid I have to tell you that I belie company operations you own; I'm afraid I've got to get some detail.' <p> ou know me?' he said grumpily. `I'm afraid I don't know you. I've shaved of that band of loonies? I'm afraid it's just no contest at all. <h hellip; we cocoon because we are afraid of saying the wrong thing or not a Pullman, and partly because he was afraid of being seen there by some attacks. They think she is afraid of upsetting the politicians lin Into the lens, but equally do not be afraid of photographing against the lig because health professionals were afraid to cross lines marking religious willowy decadence, blase, stupid, afraid to be seen sober or with a book, t say yea <p> or nay, and he was afraid to ask. <p> When the new field l with their eyes what they were afraid to put into words. Is it possib Monica) Lewinsky report, you are afraid to put a cigar in your mouth. Th
A concordance ...

s flagship Odeon cinema said: `I am afraid Ally's been subbed." <p> Rangers notions in ten years. I'm afraid I have to tell you that I belie company operations you own; I'm afraid I've got to get some detail." <p> ou know me?" he said grumpily. `I'm afraid I don't know you. I've shaved of that band of loonies? I'm afraid it's just no contest at all. <h ellip; we cocoon because we are afraid of saying the wrong thing or not a Pullman, and partly because he was afraid of being seen there by some attacks. They think she is afraid of upsetting the politicians lin Into the lens, but equally do not be afraid of photographing against the lig because health professionals were afraid to cross lines marking religious willowy decadence, blase, stupid, afraid to be seen sober or with a book, t say yea <p> or nay, and he was afraid to ask. <p> When the new field l with their eyes what they were afraid to put into words. Is it possib (Monica) Lewinsky report, you are afraid to put a cigar in your mouth. Th

.. illustrates the typical patterns of a word (collocations)
afraid /əˈfred/  

1. If you are **afraid** of someone or **afraid** to do something, you are frightened because you think that something very unpleasant is going to happen to you.  
   - She did not seem at all afraid... I was afraid of the other boys... I'm still afraid to sleep in my own bedroom.

2. If you are **afraid** for someone else, you are worried that something horrible is going to happen to them.  
   - She's afraid for her family in Somalia.

3. If you are **afraid** that something unpleasant will happen, you are worried that it may happen and you want to avoid it.  
   - I was afraid that nobody would believe me... The Government is afraid of losing the election...

4. If you want to apologize to someone or to disagree with them in a polite way, you can say... **I'm afraid**.
   - [SPOKEN] We don't have anything like that, I'm afraid... I'm afraid I can't help you... 'Bad news?' — 'I'm afraid so.'
The COBUILD project

- University of Birmingham and Collins Publisher.
- Corpus: Bank of English
- Editor John Sinclair
- 1987 first corpus dictionary
Non-literary and literary text

- General patterns / norms and deviations from norms
- Corpus work is comparison
- ‘Collocations’ and ‘semantic prosodies’
Days
What are days for?
Days are where we live.
They come, they wake us
Time and time over.
They are to be happy in:
Where can we live but days?

Ah, solving that question
Brings the priest and the doctor
In their long coats
Running over the fields

Philip Larkin
Try WebCorp Live
Corpus work is comparison

I, however, went to Rovaniemi in April when the days are long and sunny, and my face got very brown. Rovaniemi

fashion accessories in the classroom. ‘Youngsters these days are more conscious of current trends than ever before,’ said D

or rather different purposes: The pre-application visiting days are intended for students who want to find out more about the

he thinks he can push me around now. Those bad old days are gone for good. She raised her eyes to Nettie, her

cooler — and wetter, it's true! — yet the days are still long and it's easier to avoid the crowds. So

drive. Third. To understand and enjoy Versailles, several days are necessary: rather than make several trips from a Paris base

n't got the wind,’ said Vernon. ‘My kicking days are over.’ Meredith explained it was touch-line supporters the

and greed. There is a time in the day when the days are long, when the edge of light meets the edge of darkness

ion until the frost stops them’. Reference to pre-Chelsea days are rare, but Miller described in the Dictionary how he grew

a mud hut from a rainstorm in Ethiopia. Their travelling days are now all but over, following a recent trip to Guyana,
o. We're moving out of London altogether. My publicity days are over, I'm afraid.’ ‘Afraid?’ He
to ring and book in advance though because the activity days are very popular. A day of activities will cost you £12.50.
tell you that Excel has been around for too long and its days are numbered. There is truth in both views. So who are

. But if the church has its way, the garden's days are numbered. When vicar, Derek Sawyer retires, the church

as he says, ‘it all adds up’. These days are coming to an end now that he is 40 — his birthday

, shows lighting up time to be 18.52. Even though the days are still quite mild, the nights are cool, the mornings are

Elliots, which gave rise to a long Border feud. Those days are remembered in a fine statue in Galashiels of a mounted r

amending ambitions to take account of aging. My brave days are over and I now ride for relaxation rather than a fall.

beasts that walk like men. They gather now that the last days are near.’ He looked directly at Morrsleib, the lesser moon

are you on er Already [unclear] May? You're on three days are you, or what? No one day. Oh just the

voice was flat, devoid of emotion. ‘The next few days are going to be a long, hard slog. Not much time

of herring that used to arrive each July. Sadly, these days are gone, and, it seems, Man's rapacious rape of
| Corpus work is comparison                      | days are over he would like to make a new career on the turf |
| I, however, went to Rovaniemi in April when the | days are long and sunny, and my face got very brown. Rovaniemi |
| fashion accessories in the classroom. ‘Youngsters these | days are more conscious of current trends than ever before,’ said D |
| or rather different purposes: The pre-application visiting | days are intended for students who want to find out more about the |
| he thinks he can push me around now. Those bad old | days are gone |
| cooler — and wetter, it's true! — yet the | days are still |
| drive. Third. To understand and enjoy Versailles, several | days are new |
| n't got the wind,’ said Vernon. ‘My kicking | days are over |
| and greed. There is a time in the day when the | days are long |
| modigation until the frost stops them’. Reference to pre-Chelsea | days are rare, but Miller described in the Dictionary how he grew |
| a mud hut from a rainstorm in Ethiopia. Their travelling | days are now all but over, following a recent trip to Guyana, |
| so. We're moving out of London altogether. My publicity | days are over, I'm afraid.’ ‘Afraid?’ He |
| to ring and book in advance though because the activity | days are very popular. A day of activities will cost you £12.50. |
| tell you that Excel has been around for too long and its | days are numbered. There is truth in both views. So who are |
| , shows lighting up time to be 18.52. Even though the | days are coming to an end now that he is 40 — his birthday |
| Elliot, which gave rise to a long Border feud. Those | days are still quite mild, the nights are cool, the mornings are |
| amending ambitions to take account of aging. My brave | days are remembered in a fine statue in Galashiels of a mounted r |
| beasts that walk like men. They gather now that the last | days are over and I now ride for relaxation rather than a fall. |
| are you on er Already [unclear] May? You're on three | days are near.’ He looked directly at Morrsleib, the lesser moon |
| voice was flat, devoid of emotion. ‘The next few | days are you, or what? No one day. Oh just the |
| of herring that used to arrive each July. Sadly, these | days are going to be a long, hard slog. Not much time |
| days are gone, and, it seems, Man's rapacious rape of |
Opportunistic or mainstream?

"[i]t is a truism that a corpus is neither good nor bad in itself, but suited or not suited to a particular purpose."

(Hunston 2002: 26)
Representativeness

the extent to which the results of a corpus analysis allow for generalisations to be made that apply to the language as a whole (or at least to one language variety).
BNC 100m words

http://www.natcorp.ox.ac.uk/what/index.html

- a **sample** corpus: composed of text samples generally no longer than 45,000 words.
- a **synchronic** corpus: the corpus includes imaginative texts from 1960, informative texts from 1975.
- a **general** corpus: not specifically restricted to any particular subject field, register or genre.
- a **monolingual** British English corpus: it comprises text samples which are substantially the product of speakers of British English.
- a **mixed** corpus: it contains examples of both spoken and written language.
### KEYWORD IN CONTEXT DISPLAY

<table>
<thead>
<tr>
<th>KEYWORD</th>
<th>CONTEXT</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAYS ARE</td>
<td></td>
<td>276</td>
</tr>
</tbody>
</table>

#### CLICK FOR MORE CONTEXT

1. **G3J W_fct_prose**
   - *house licence or 1 refreshment licence, the hour's for week*

2. **AHX W_newsp_bright_net_mis**
   - *tickets which offer unlimited travel on British Rail for a fee*

3. **CJR W_pop_liter**
   - *management skills -- both political and organisational. The point*

4. **CCC W_religion**
   - *because you can see nothing in it. The old*

5. **BYI W_biography**
   - *days of your life, but I do think that student*

6. **BMF W_pop_liter**
   - *and authorities that represent their fellow users. Practice*

7. **CRS W_sc_pol_law.edu**
   - *and adjusting to its demands and possibilities. The system's*

8. **AKS W_newsp_bright_net_mis**
   - *the Mercantile Credit in Bournemouth last January. Those dark*

9. **GSL W_consult**
   - *it's we it's it's (unclear) I mean some*

10. **ARU W_pop_liter**
    - *at work. Either way, these next two or three*

11. **HJR W_fct_prose**
    - *are the toys of the telephone. Do not think these*

12. **AKM W_newsp_bright_net_mis**
    - *per cent of all new cars sold in Western Europe these*

13. **HTE W_misc**
    - *friends; teachers and careers staff; Departmental Visiting*

14. **KCO W_conv**
    - *diary. It's all in, I mean (pause) the*

15. **K73 W_courtroom**
    - *Lordship has not found that the last five or a half*

16. **K2G W_newsp_other_science**
    - *some pot pourri. Memories of good old Andy and beachcomber*

17. **HOT W_non_ac_soc_sci**
    - *(LG: Parents in need. Many families these*

18. **FP0 W_fct_prose**
    - *they said, and in her dissertation, Chomsky thought*

19. **HSH W_pop_prose**
    - *or planting bed. A Centre of Attraction. Just because*

20. **HTI W_fct_prose**
    - *is still young, 'he warned. Though the*

21. **GUI W_fct_prose**
    - *was flat, devoid of emotion. The next few*

22. **KRU W_bbcast_news**
    - *the first test of his promise that the bad old*

23. **B1N W_misc**
    - *au 'balancing very much to history. Those having fishing*

24. **MTN W_fct_prose**
    - *when we were children. They were happy. But those*

25. **H24 W_newsp_script**
    - *at Rover's body plant, and saw the bad old*

26. **BNH W_misc**
    - *having fun are just some of the reasons why corporate selling*

27. **CBN W_religion**
    - *life. 'Well it's true enough that some school*
Types of corpora

- General Purpose Corpus
- Monitor Corpus
- Sample Corpus
- Mini-corpus
- Multilingual Corpus
- Comparable Corpus
- Parallel Corpus
- Translation Corpus
- Diachronic Corpus
- Learner Corpus
... for the increase. I see an Hon. Member from Alberta nodding his head in agreement. All that means is that once the rates go up and fa...

... 60> Instead, we saw the Minister of Transport sitting across the aisle with his hands too busy worrying about this. He was too busy to wo ...

... ng of the Hon. Member for St. John's East. I see him nodding his head in the affirmative, and I am sure that he is very pleased with that initiative. ...

... I see his colleague, the Hon. Member for Moose Jaw nodding his head in full agreement. At one time the Farm Credit Corporation had 6 ...

... to any kind of action. There seems to be a dichotomy between his head Knowledge and his ability to respond to it. I wish to draw the a ...

... ents would cure the problems. I see someone opposite shaking his head. Let me tell him that I want to deal with one or two more points ...

... certain forms of change to the Crow. The Hon. Member shakes his head. Mr. Speaker, I cannot remember exactly what Member asked the qu...

... id=2044856>1 I do not believe the Minister even said that and he is shaking his head negatively. However that feeling is obviously out there across t ...

... aranteed profits to CP. The Minister of Transport is shaking his head. No. Let me deal with my area of moment, for a ...

... he Hon. Member for Mississauga North (Mr. Fisher) was very anxious to shake his head. That is not a point of order. There does not app ...

... domain. A Crown corporation and the political appointment of its head are matters that are certainly within the public domain. I am go ...

... the block, if that is the right place. Well, it should have its head chopped off, having a deficit of $1.5 billion. It claims to be r ...

... Our recent customer, of course, is Argentina. Its head of state is on record as publicly saying that he wants the Candu reactor pro...

... olution in communications technology would make it possible for firms to locate head offices away from the largest population centres, the Government guts the regio ...

... the Tory side of the House over to the Liberal side and are automatically made head of a Crown corporation so that they can tell the Tories where to go on railway ...

... 059145> He is prepared to take on Members from Quebec and/or Government Members head to head. At least there will be a voice from Quebec in the House ...

... s of high finance, economics and so forth, but I still cannot get it through my head what kind of economic sense there is in what the Government is trying to prom ...

... 35430> I question whether or not this is the time to get into a $30 million new head office. I question the misuse of the resources of Air Canada, wh ...

... he Saskatchewan border. Those four huge ranches ran thousands of head of cattle. When the Minister recognizes the history of that Prov ...

... has seen fit to bring in this Bill at the eleventh hour, holding the gun to our head as Parliamentarians and forcing us to deal with all phases of the Bill in one d ...

... a socialist country. I can remember travelling with the previous head of Canadian Pacific Railways and the head of the CNR. We were vi ...

... > For that kind of money, these people would have a substantial head start in producing railway axles, rolling stock, modern marine fittings or some ...

... inster of the Environment, but I do not think his response hit the nail on the head. He tried to fudge around on the issue, saying that UIC Section ...

... ant to let the railways take it. I invite you to kick me in the head, Mr. Speaker— Hear, hear!—but I should like to ...

... ssay. There is atemptation for Ministers here in Ottawa at the head of large Departments to think that they know best. We are very m ...

... at back in 1975. Where is he now? He is the head of the CBS and is now rewriting the Broadcasting Act. Culture an ...

... id=2033631> Madame Speaker, I have a supplementary question. The head of the Argentinian Atomic Agency was quoted at the opening of the Candu reactor ...

... indeed. May I also say that I am meeting with David Ahenakew, the head of the AFN, here in Ottawa this afternoon and we may adopt a type of umbrella a ...

... Some of the people on the trip said afterwards: «You know, the head of CP sounded more like he was running a Crown corporation than the head of the oth ...

... to be the head of Canagrex although I know that I am fully capable of being the head of Canagrex. I know there are other people who are just as capab ...

... remember travelling with the previous head of Canadian Pacific Railways and the head of the CNR. We were visiting a proposed construction site betwee ...

... he head of CP sounded more like he was running a Crown corporation than the head of the other people. They thought the heads of the two railroads, by their ph ...

... two Canadians, with the people of Alert Bay and Sointula and the one at the head of Johnstone Strait that demand absolutely, absolutely, and unequivocally from the
Size

- 1 million words was “big” in the 1960s-1980s but now is nothing special. BNC = 100m of which 90m written
- the bigger the better - or small is beautiful?
- to study frequent items you can manage with a smaller corpus
- Size and text length
Documentation

- Standard header (at top of each text or in separate file/website)
- POS, errors/non-standard forms, text structure
Practical constraints

- Typing in v. Scanning v. Finding
- Recording
- Using what is out there?
Legal and moral constraints

- “Anonymizing” texts e.g. by replacing names with NAME to preserve confidentiality
- Requesting copyright permission
- Impossibility of getting e.g. business negotiations.
- Ethics procedures (very important for assignments already!!)
Corpus design: example sustainable development corpus

- Cultural key word
- Discourse as representation
- World Summit, 26 August to 4 September 2002 in Johannesburg
- *Guardian* 2002
  - 368 occurrences of SD
  - 211 articles
  - approximately 150,000 words

What Corpus Tools can do

- create word lists sorted in various ways
- find examples of words or phrases you want to seek out
- sort these examples in numerous ways for you
- process your texts in pre-determined ways, e.g.
  - parse them in Part of Speech (POS) categories
  - identify key words (KWs)
  - show you where each word or phrase comes in the source text
  - separate out words you’re not interested in
  - plot the history of word use
What Corpus Tools can’t do

- suggest a good research question
- supply answers to your research questions
- understand the texts
- E.g. find examples of metaphors or complaining or swearing in your texts for you
Some conclusions

- Corpus linguistics provides methods that can be used in almost any area of language study
- Advantages of corpus methods: objectivity, replicability, more than intuition
- New terminology: e.g. ‘concordance’, ‘KWIC’, ‘collocations’, ‘semantic prosodies’
- Go and have a try!
Introductory reading


Download Open Access articles:
http://www.nottingham.ac.uk/english/people/michaela.mahlberg
The International Journal of Corpus Linguistics (IJCL) publishes original research covering methodological, applied and theoretical work in any area of corpus linguistics. Through its focus on empirical language research, IJCL provides a forum for the presentation of new findings and innovative approaches in any area of linguistics (e.g. lexicology, grammar, discourse analysis, stylistics, sociolinguistics, morphology, contrastive linguistics), applied linguistics (e.g. language teaching, forensic linguistics), and translation studies. Based on its interest in corpus methodology, IJCL also invites contributions on the interface between corpus and computational linguistics. The journal has a major reviews section publishing book reviews as well as corpus and software reviews. The language of the journal is English, but contributions are also invited on studies of languages other than English. IJCL occasionally publishes special issues (for details please contact the editor). All contributions are peer-reviewed.

IJCL offers authors the option to publish articles as Open Access, click here for an example.
Have a look at

- corpora list
  - http://www.hit.uib.no/corpora/sub.html
Bookmarks for Corpus-based Linguists

David Lee

* WELCOME *

* The official URL for this site has been changed to http://tiny.cc/corpora

Please update your bookmarks/favourites/links.

The "Fast URL" I've been using for many years ("http://devoted.to/corpora") can no longer be used.

* Sorry for this unexpected change! *

Why create this site? Well, because almost all similar sites are really outdated (full of dead links) and haven't been spring-cleaned for years. In addition, they are not as comprehensive or wide-ranging in scope, or don't cater to applied linguists. Having this bookmarks/favourites site on the web means all you corpus-based linguists out there won't have to each create your own for your individual computers. (So... if you find a really good web site or resource not already listed here, don't keep it to yourself! Let me know, and I'll put it up here to share with others.) It also means we don't all have to individually remove dead links or update our links on our separate web pages. Let's make this web site a collaborative resource.

What's here? These annotated links (c. 1,000 of them) are meant mainly for linguists and language teachers who work with corpora, not computational linguists/NLP (natural language processing) people, although the latter are certainly welcome to find and use them as well. A selection (for