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 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 for using tissue at the
Pushing envelopes Pushing envelopes: red nose has to pay through the nose for insurance' Do you live
Welsh are paying through the 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 job? He's set to wed
Does Mike Tindall need a 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
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 The owner of two barking
Great Dane bites off courier's nose The Guardian reports on a
1837: Woman bites off man's nose
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 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 Japan Elderly woman breaks student's nose for the unusual By Mike 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 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 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 vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked prevents sniffles Having a 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 vigorous exercise at 7.9% and nose 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 day Pride and embarrassment has to pay 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 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 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 vigorously 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
a good bargain on 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
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 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 operation could prove
Ed Nose Day' Ed Miliband's nose vigilante Phoenix Jones suffers broken nose Phoenix Jones, the masked
Science News Having a big nose prevents sniffles Having a big nose
Sunday | Home More on: runny nose The owner of two barking
Great Dane bites off courier's nose The Guardian reports on a
1837: Woman bites off man's nose
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
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 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 vigorous exercise 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 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 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
t 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 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 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:

Search API: Google

Language:

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

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
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 believ
compant operations you own; I'm afraid I've got to get some detail.' <p>
o 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
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 believ
company operations you own; I'm afraid I've got to get some detail.' <c
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
cellip; 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 /əfriːd/

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. 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
<table>
<thead>
<tr>
<th>Corpus work is comparison</th>
<th>days are over he would like to make a new career on the turf</th>
</tr>
</thead>
<tbody>
<tr>
<td>. I, however, went to Rovaniemi in April when the</td>
<td>days are long and sunny, and my face got very brown. Rovaniemi</td>
</tr>
<tr>
<td>fashion accessories in the classroom. ‘Youngsters these</td>
<td>days are more conscious of current trends than ever before,’ said D</td>
</tr>
<tr>
<td>or rather different purposes: The pre-application visiting</td>
<td>days are intended for students who want to find out more about the</td>
</tr>
<tr>
<td>he thinks he can push me around now. Those bad old</td>
<td>days are gone for good. She raised her eyes to Nettie, her</td>
</tr>
<tr>
<td>cooler — and wetter, it's true! — yet the</td>
<td>days are still long and it's easier to avoid the crowds. So</td>
</tr>
<tr>
<td>drive. Third. To understand and enjoy Versailles, several</td>
<td>days are necessary: rather than make several trips from a Paris base</td>
</tr>
<tr>
<td>n't got the wind,’ said Vernon. ‘My kicking</td>
<td>days are over.’ Meredith explained it was touch-line supporters the</td>
</tr>
<tr>
<td>and greed. There is a time in the day when the</td>
<td>days are long, when the edge of light meets the edge of darkness</td>
</tr>
<tr>
<td>ion until the frost stops them’. Reference to pre-Chelsea</td>
<td>days are rare, but Miller described in the Dictionary how he grew</td>
</tr>
<tr>
<td>a mud hut from a rainstorm in Ethiopia. Their travelling</td>
<td>days are now all but over, following a recent trip to Guyana,</td>
</tr>
<tr>
<td>o. We're moving out of London altogether. My publicity</td>
<td>days are over, I'm afraid.’ ‘Afraid?’ He</td>
</tr>
<tr>
<td>to ring and book in advance though because the activity</td>
<td>days are very popular. A day of activities will cost you £12.50.</td>
</tr>
<tr>
<td>tell you that Excel has been around for too long and its</td>
<td>days are numbered. There is truth in both views. So who are</td>
</tr>
<tr>
<td>. But if the church has its way, the garden's</td>
<td>days are numbered. When vicar, Derek Sawyer retires, the church</td>
</tr>
<tr>
<td>as he says, ‘it all adds up’. These</td>
<td>days are coming to an end now that he is 40 — his birthday</td>
</tr>
<tr>
<td>, shows lighting up time to be 18.52. Even though the</td>
<td>days are still quite mild, the nights are cool, the mornings are</td>
</tr>
<tr>
<td>Elliots, which gave rise to a long Border feud. Those</td>
<td>days are remembered in a fine statue in Galashiels of a mounted r</td>
</tr>
<tr>
<td>amending ambitions to take account of aging. My brave</td>
<td>days are over and I now ride for relaxation rather than a fall.</td>
</tr>
<tr>
<td>beasts that walk like men. They gather now that the last</td>
<td>days are near.’ He looked directly at Morrsleib, the lesser moon</td>
</tr>
<tr>
<td>are you on er [unclear] May? You're on three</td>
<td>days are you, or what? No one day. Oh just the</td>
</tr>
<tr>
<td>voice was flat, devoid of emotion. ‘The next few</td>
<td>days are going to be a long, hard slog. Not much time</td>
</tr>
<tr>
<td>of herring that used to arrive each July. Sadly, these</td>
<td>days are gone, and, it seems, Man's rapacious rape of</td>
</tr>
</tbody>
</table>
Corpus work is comparison gone / numbered / over
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.
<table>
<thead>
<tr>
<th>267 Tokens</th>
<th>Banking, Travel, and Communication Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>G31 W_fict_prose A B C</td>
</tr>
<tr>
<td>2.</td>
<td>AHX W_newsp_brith_net_misc A B C</td>
</tr>
<tr>
<td>3.</td>
<td>CJR W_pop_lore A B C</td>
</tr>
<tr>
<td>4.</td>
<td>CCC W_religion A D C</td>
</tr>
<tr>
<td>5.</td>
<td>BLY W_biology A D C</td>
</tr>
<tr>
<td>6.</td>
<td>BMP W_pop_lore A D C</td>
</tr>
<tr>
<td>7.</td>
<td>CRS W_ac_polit_law_aedu A D C</td>
</tr>
<tr>
<td>8.</td>
<td>AKS W_newsp_brith_net_misc A D C</td>
</tr>
<tr>
<td>9.</td>
<td>GML S_consult A D C</td>
</tr>
<tr>
<td>10.</td>
<td>ARU W_pop_lore A B C</td>
</tr>
<tr>
<td>11.</td>
<td>HGR W_fict_prose A B C</td>
</tr>
<tr>
<td>12.</td>
<td>AKM W_newsp_brith_net_misc A B C</td>
</tr>
<tr>
<td>13.</td>
<td>HTE W_misc A B C</td>
</tr>
<tr>
<td>14.</td>
<td>KEC S_cwav A B C</td>
</tr>
<tr>
<td>15.</td>
<td>K73 S_courtroom A C</td>
</tr>
<tr>
<td>16.</td>
<td>KSC W_newsp_other_science A B C</td>
</tr>
<tr>
<td>17.</td>
<td>HOT W_non_ac_soc_society A B C</td>
</tr>
<tr>
<td>18.</td>
<td>FP0 W_fict_prose A B C</td>
</tr>
<tr>
<td>19.</td>
<td>HSX W_pop_prose A B C</td>
</tr>
<tr>
<td>20.</td>
<td>H7W W_fict_prose A B C</td>
</tr>
<tr>
<td>21.</td>
<td>GUT W_fict_prose A B C</td>
</tr>
<tr>
<td>22.</td>
<td>KRU C_broadcast_news A B C</td>
</tr>
<tr>
<td>23.</td>
<td>B1N W_misc A B C</td>
</tr>
<tr>
<td>24.</td>
<td>HTN W_fict_prose A B C</td>
</tr>
<tr>
<td>25.</td>
<td>K24 W_news_script A B C</td>
</tr>
<tr>
<td>26.</td>
<td>BNH W_misc A B C</td>
</tr>
<tr>
<td>27.</td>
<td>CBN W_religion A B C</td>
</tr>
<tr>
<td>28.</td>
<td>CES W_fict_prose A B C</td>
</tr>
</tbody>
</table>
Types of corpora

- General Purpose Corpus
- Monitor Corpus
- Sample Corpus
- Mini-corpus
- Multilingual Corpus
- Comparable Corpus
- Parallel Corpus
- Translation Corpus
- Diachronic Corpus
- Learner Corpus
Vous ne serez probablement pas d'accord, monsieur le Président...
Les ministres qui dirigent ici, à Ottawa, de grands ministères sont portés à croire qu'eux seuls ont toujours raison.

Il est président de Radio-Canada et en ce moment il écrit la loi sur la radiodiffusion.

Le chef de la Commission atomique de l'Argentine aurait bel et bien dit lors de l'inauguration du réacteur Candu que les négociations avec la Libye étaient en cours.

Je tiens à ajouter que je rencontre David Ahenakew, chef de l'APN à Ottawa cet après-midi et nous essayerons peut-être d'adopter une entente globale et nationale, calquée sur l'entente avec le Manitoba.

Certains de nos compagnons de voyage nous ont dit par la suite : « Vous savez, le président du CP se comporte beaucoup plus comme s'il dirigeait une société de la Couronne que le président du CN ».

Je n'ai nullement envie de devenir chef de cette société, même si je suis parfaitement capable, j'en suis persuadé, d'occuper ce poste.

Je me souviens d'un voyage que j'ai fait avec l'ex-président du Canadien Pacifique et le président du CN.

Certains de nos compagnons de voyage nous ont dit par la suite : « Vous savez, le président du CP se comporte beaucoup plus comme s'il dirigeait une société de la Couronne que le président du CN ».

Les activités de pêche dans le détroit de Johnstone sont parmi les plus compliquées de l'ouest du Canada, néanmoins la population d'Alert Bay, de Sointula et de la région en amont du détroit de Johnstone, Alert Bay est une petite localité insulaire située à la tête du détroit de Johnstone.

Je n'ai nullement envie de devenir chef de cette société, même si je suis parfaitement capable, j'en suis persuadé, d'occuper ce poste.

Le système américain est légèrement différent du système canadien et du système britannique en ce sens que le chef d'État, le président, est élu par un scrutin populaire indépendant et qu'il n'accède pas de manière permanente. Les civils qui en doutent n'ont qu'à s'adresser au directeur de l'Association canadienne du football amateur.

Le comité du cabinet est dirigé par Honoré Don Phillips.

Le directeur de la Banque de Cologne d'Allemagne de l'Ouest était récemment de passage dans ma ville.

2033631: Madame Speaker, I have a supplementary question. The head of the Argentine Atomic Agency was quoted at the opening of the Candu reactor...
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
International Journal of Corpus Linguistics

Editor Michaela Mahlberg | Nottingham, UK
Consulting Editor Wolfgang Teubert | Birmingham, UK
Reviews Editor Beatrix Busse | Heidelberg, Germany

ISSN 1384-6655 | E-ISSN 1569-9811
Electronic edition at ingentaConnect

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](http://www.hit.uib.no/corpora/sub.html)

Bookmarks for Corpus-based Linguists

* 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 may find some of the links useful as well.