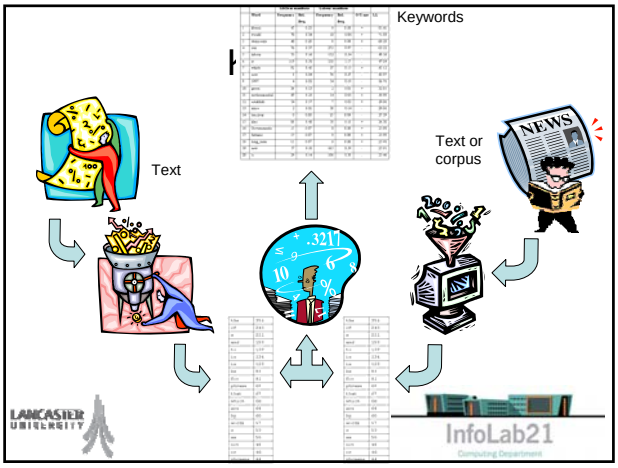
# The key domain method for the study of language varieties

Paul Rayson and Nicholas Smith
UCREL, Computing Department,
Lancaster University, UK.

## Outline

- New kind of method and tool (Matrix) for the statistical analysis of corpora
- Matrix integrates part-of-speech tagging and semantic field tagging in a profiling tool
- Extends the keywords procedure to identify key grammatical categories and key concepts
- Case studies:
  - Key concepts in 20th C. romantic fiction
  - 'Obligation & necessity' in 20th C. corpora
- Software demo



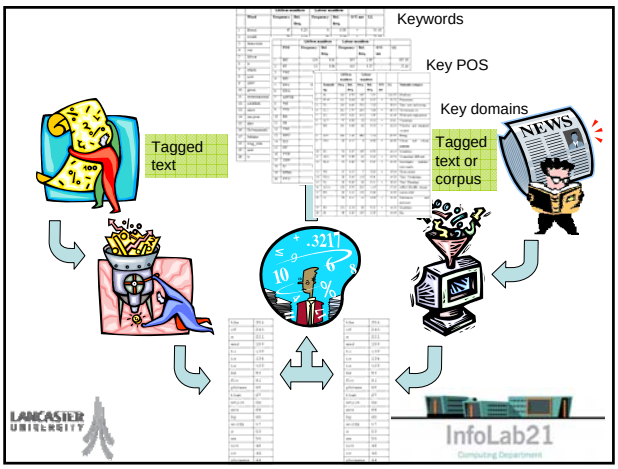

## Log likelihood statistic

| | Corpus one | Corpus two | Total |
|---|---|---|---|
| **Frequency of word** | a | b | a+b |
| **Frequency of word not occurring** | c-a | d-b | c+d-a-b |
| **TOTAL** | c | d | c+d |

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

$$-2\ln\lambda = 2\sum_i O_i \ln\left(\frac{O_i}{E_i}\right)$$

E1 = c × (a+b) / (c+d)
E2 = d × (a+b) / (c+d)
LL = 2×((a×ln (a/E1)) + (b×ln (b/E2)))




## Matrix method

- Integrates POS tagging and semantic field annotation into a profiling tool
- Extends keywords procedure to identify key grammatical categories and key concepts

## Previous studies using Matrix

- Social differentiation in the use of English vocabulary (Rayson, Leech & Hodges, 1997)
- Profiling of learner English (Granger & Rayson, 1998)
- Semantic analysis of technical documents from the software engineering domain (Sawyer, Rayson & Garside, 2005)

## The data for today's case studies

- Standard written British English
- Sampling dates 1931 – 1961 – 1991
- Each corpus contains 1M words
- 15 genres of published informative and imaginative prose (e.g. press reportage, academic writing, romantic fiction, science fiction)
- Corpus size and sampling frame modelled on the Brown Corpus (of 1960s American English)

## Comparable corpora across the C20th

| | **1901 (1898-1904)** | **1931 (1928-34)** | **1961** | **1991/ 1992** |
|---|---|---|---|---|
| **BrE** | Lanc-1901 | **B-LOB (Lanc-1931)** | **LOB** | **F-LOB** |
| AmE | ? | Pre-Brown31 | Brown | Frown |

## Example 1: A bottom-up view of Romantic fiction across the 20th C.

General questions:

- What 'key concepts' help to distinguish romantic fiction from fiction in general?
- What changes among 'key concepts' have changed over the course of the 20th century?

## Romantic fiction vs. Fiction in general

- Synchronic comparison between the F-LOB corpus and the BNC-Sampler Imaginative subcorpus
- Intuitive expectations about key concepts: LOVE, RELATIONSHIPS, …

**Items used significantly more frequently in Romantic fiction than in general fiction (FLOB vs. BNC-Sampler Imaginative subcorpus)**

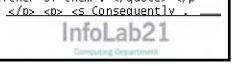
| Tag | Gloss | FLOB romantic | | Sampler Imag | | Over/ under use | LL |
|---|---|---|---|---|---|---|---|
| | | freq | % | freq | % | | |
| A5.3+ | Evaluation:- Accuracy | 125 | 0.2 | 150 | 0.1 | + | 69.0 |
| T1.1 | Time: General | 27 | 0.0 | 4 | 0.0 | + | 58.3 |
| L1+ | Life and living things | 47 | 0.1 | 38 | 0.0 | + | 42.1 |
| X2.6- | Expect | 40 | 0.1 | 28 | 0.0 | + | 40.9 |
| B1 | Anatomy / physiology | 1104 | 1.7 | 3057 | 1.4 | + | 33.6 |
| X2.6+ | Expect | 74 | 0.1 | 101 | 0.1 | + | 33.1 |
| X2.4 | Investigate, examine, test, search | 128 | 0.2 | 233 | 0.1 | + | 30.1 |
| S1.1.1 | Social actions, states and processes | 127 | 0.2 | 237 | 0.1 | + | 27.8 |
| T1.3- | Time: Period | 58 | 0.1 | 77 | 0.0 | + | 27.2 |
| O4.2+ | Judgement of appearance (pretty etc.) | 194 | 0.3 | 448 | 0.2 | + | 19.3 |
| N3.7+ | Measurement: Length & height | 86 | 0.1 | 164 | 0.1 | + | 17.7 |
| O4.1 | General appearance and physical properities | 142 | 0.2 | 322 | 0.1 | + | 15.4 |
| A12++ | Easy/difficult | 11 | 0.0 | 7 | 0.0 | + | 12.2 |
| X5.1- | Attention | 25 | 0.0 | 34 | 0.0 | + | 11.2 |

## Key tag ‘EXPECT’ in Romantic fiction (1990s BrE)

| | | |
|---|---|---|
| Her attitude took Frances by | surprise | , for it was in such marked c |
| around him , catching him by | surprise | - which might not have matter |
| most of Father 's unexpected | surprises | , in which the gift was of su |
| yes . <s There was something | unpredictable | in her that was only explaine |
| er nipple as she gasped with | surprise | , sucking so it puckered poin |
| y , when their future seemed | hopeless | , Judy had an idea .   |
| rs .   <s <quote> How | amazingly | quiet it is , </quote> she sa |
| e you ever known anything so | amazingly | quiet ? </quote>   <s |
|   <s <quote> I 'm not | surprised | , </quote> said Tom .  <p |
| discoveries . <s It would be | surprising | if there were . <s I was hopi |
| Abby turned to look at him , | surprised | at her own sense of relief at |
| ne .   <s To Jenny 's | surprise | and - astonishingly enough - |
| s To Jenny 's surprise and - | astonishingly | enough - relief , he appeared |
| f her could n't help feeling | amazed | that he had taken her tentati |
| nger . <s She was constantly | surprised | by the sensitive perception o |
| s that , as Thomas aged with | astonishing | rapidity , Sophie became even |
| <s The driver turned in some | surprise | , then laughed to see a maid |
| o the driver , he spoke with | surprising | authority .   <s <quot |
| e to you often . <s But I wo | n't hold my breath | waiting for a letter from you |
|   <s To her complete | amazement | , she received long letters b |

## Key tag ‘ATTENTION’ in Romantic fiction (1990s BrE)

| | | |
|---|---|---|
| ad had their orders and they | ignored | her , gazing intently into th |
| >  <s His younger brother | ignored | the proffered hand . <s <quot |
| low Joscelyn 's behaviour to | upset | him .   <s Yet when he |
| use and because he wanted to | distract | his mind he went immediately |
| <quote> Did you feel she 's | disturbed | ? </quote> he asked earnestly |
| e castle walls .   <s | Ignoring | the ancient tower and passing |
| proving quite devastatingly | distracting | .   <s She returned to |
| .   <s <quote> Do n't | upset | your machine , </quote> Jim a |
| make sure she was n't living | in a dream | . <s The transformation in Ra |
| art I pushed open the door , | ignoring | the man who was saying , <quo |
| put off asking Rob what was | troubling | him .   <s Sometimes , |
| Bob 's voice broke into her | reverie | , for she had n't heard him c |
| With Danny she would n't sit | aimlessly | at home with an embroidery fr |
| om Foley was asleep and deep | in a dream | in which he had just resigned |
| t people get pleasure out of | daydreams | ? </quote>   <s <quote |
| ccusing him - and he had not | troubled | to deny it . <s Why ?  <p |
| few days , and since she was | disturbing | Paul with her restlessness , |
| ge . <s Looking for some new | diversion | for the children 's amusement |
| you had someone with you to | take your mind off | business . </quote>   |
| <s <quote> But that did n't | take my mind off | you . </quote>   <s <q |
| they had finished and then , | ignoring | their studied attentiveness a |
| nd it . <s I do n't want to | upset | either of them . </quote> </p |
| ings that they were n't to be | disturbed | .   <s Consequently , |



## Diachronic development of key concepts in Romantic fiction (1): Items used significantly more frequently in 1991 than in 1961

| Tag | Gloss | FLOB imag | | Sampler Imag | | Over/ under use | LL |
|---|---|---|---|---|---|---|---|
| | | freq | % | freq | % | | |
| B1 | Anatomy and physiology | 1104 | 1.7 | 756 | 1.3 | + | 24.7 |
| B3 | Medicines and medical treatment | 131 | 0.2 | 53 | 0.1 | + | 23.8 |
| *K4* | *Drama, the theatre and showbusiness* | *88* | *0.1* | *25* | *0.0* | + | *28.6* |
| *X3.2+* | *Sensory:- Sound* | *16* | *0.0* | *3* | *0.0* | + | *8.0* |
| *S1.2.5+* | *Toughness; strong/weak* | *25* | *0.0* | *8* | *0.0* | + | *6.9* |

## Key concept ‘ANATOMY/PHYSIOLOGY’ in 1991 (FLOB)

| | | |
|---|---|---|
| shattered . <s She could n't | sleep | without Thomas in their big o |
| s appeared around her lovely | eyes | and she started having midnig |
| quote>   <s Sophie 's | eyes | filled with tears . <s <quote |
| s Sophie 's eyes filled with | tears | . <s <quote> Oh , Thomas , yo |
| in ten years his overworked | digestive tract | was behaving itself . <s He p |
| elf . <s He patted Sophie 's | cheek | with a lean brown hand . <s < |
| e 's cheek with a lean brown | hand | . <s <quote> Of course I want |
| for Sophie . <s His trousers | stayed up | only by willpower and a tight |
| t belt . <s His sun-bleached | hair | was dazzling and his curly ye |
| azzling and his curly yellow | beard | showed off his tan .   |
| yellow beard showed off his | tan | .   <s <quote> Well , |
| s Sophie arrived , plump and | breathless | , her cheeks slightly spotty |
| , plump and breathless , her | cheeks | slightly spotty . <s There ha |
| s There had been no time for | hair | appointments , and the glossy |
| thinking up new slogans , he | goes to bed | for a few days on a low-starc |
| lso delays the return of his | paunch | . <s He 's kept the beard , t |
| s paunch . <s He 's kept the | beard | , though . <s He says it 's g |
| ave won more than one lad 's | heart |   <s THE CHAPEL outing |
| d that her son Jacob had his | eyes | on the Perkins girl , Ann . < |

## Diachronic development of key concepts in Romantic fiction (2): Items used significantly more frequently in 1961 (LOB) than in 1931(B-LOB)

| Tag | Gloss | FLOB 1961 | | B-LOB 1931 | | Over/ under use | LL |
|---|---|---|---|---|---|---|---|
| | | freq | % | freq | % | | |
| T1.1.3 | Time: General: Future | 463 | 0.8 | 346 | 0.6 | + | 24.9 |
| G2.1- | Crime, law and order: | 54 | 0.1 | 25 | 0.0 | + | 12.7 |
| H2 | Parts of buildings | 245 | 0.4 | 186 | 0.3 | + | 12.1 |
| G2.1 | Crime, law and order: | 50 | 0.1 | 23 | 0.0 | + | 11.9 |
| X2.6+ | Expect | 86 | 0.2 | 51 | 0.1 | + | 11.3 |
| *Z1* | *Personal names* | *1506* | *2.7* | *1117* | *1.9* | + | *84.1* |
| *Z3* | *Other proper names* | *117* | *0.2* | *71* | *0.1* | + | *14.4* |
| *M1* | *Moving, coming and going* | *909* | *1.61* | *841* | *1.4* | + | *8.42* |
| *W2* | *Light* | *73* | *0.13* | *46* | *0.08* | + | *7.94* |
| *A7+* | *Definite (+ modal(s)* | *705* | *1.25* | *649* | *1.08* | + | *6.99* |
| *A14* | *Exclusivizers/particularizers* | *254* | *0.45* | *212* | *0.35* | + | *6.79* |
| *Q2.2* | *Speech acts* | *665* | *1.18* | *611* | *1.02* | + | *6.77* |
| *Z8* | *Pronouns etc.* | *7436* | *13.16* | *7577* | *12.62* | + | *6.68* |

## Key concept ‘TIME: GENERAL: FUTURE’ in 1961 (LOB) vs. 1931 (B-LOB)

| | | |
|---|---|---|
| . <quote> ooh , it 's late - I | 'll | have to be going home for my d |
| going home for my dinner or I | 'll | be late back for school </quot |
| Tommy 's direction . <quote> I | 'll | see you across the road &helli |
| 'cos Grandma 's ill but she 's | going to | get a job soon </quote> . he p |
| but she 's going to get a job | soon | </quote> . he paused to consid |
| the Town Hall clock when it 's | going to | strike </quote> . he burred se |
| im across the road . <quote> I | 'll | take you home , </quote> he co |
| she did n't mind , in fact the | sooner | the better and more luck next |
| ooner the better and more luck | next time | was her motto . she had been t |
| <quote> oh that - I expect you | 'll | be thinking of having one like |
| te> one of your many young men | will | be sweeping you off your feet |
| on . <quote> come along and we | 'll | have a cup of tea before I tak |
| from Glasgow to talk over her | future | with Mr Robertson and the loca |
| pson sounded anxious . <quote> | will | n't you be lonely ? </quote> < |
| is n't it kind of her ? I - I | 'll | never forget all she did for m |
| fit and - </quote> <quote> you | will | need a full purse ! </quote> M |
| ave lunch with her at her flat | next day | . then she followed the maid u |
| e> she explained . <quote> she | 'll | soon make you feel at home . < |
| he explained . <quote> she 'll | soon | make you feel at home . </quot |

## Diachronic development of key concepts in Romantic fiction (3): Items used significantly more frequently in 1931 (B-LOB) than in 1961(LOB)

| Tag | Gloss | B-LOB 1931 freq | % | LOB 1961 freq | % | Over/ under use | LL |
|---|---|---|---|---|---|---|---|
| A13.5 | Degree: Compromisers | 104 | 0.2 | 49 | 0.1 | + | 17.0 |
| A13.2 | Degree: Maximizers | 124 | 0.2 | 64 | 0.1 | + | 16.0 |
| E4.1+++ | Happy/sad: Happy | 17 | 0.0 | 2 | 0.0 | + | 12.7 |
| N3.2- | Measurement: Size | 156 | 0.3 | 95 | 0.2 | + | 11.5 |
| S7.1- | Power, organizing | 46 | 0.1 | 18 | 0.0 | + | 11.0 |
| *G1.2* | *Politics* | *20* | *0.0* | *2* | *0.0* | + | *16.0* |
| *A1* | *GENERAL AND ABSTRACT TERMS* | *8* | *0.0* | *0* | *0.0* | + | *10.6* |
| *N1* | *Numbers* | *341* | *0.6* | *245* | *0.4* | + | *10.5* |
| *A5.1+* | *Evaluation:- Good/bad* | *169* | *0.3* | *108* | *0.2* | + | *10.1* |
| *A13.3* | *Degree: Boosters* | *417* | *0.7* | *313* | *0.6* | + | *9.2* |
| *A13.7* | *Degree: Minimizers* | *65* | *0.1* | *33* | *0.1* | + | *8.8* |
| *T3+* | *Time: Old, new and young; age* | *176* | *0.3* | *117* | *0.2* | + | *8.6* |

## Key concept 'DEGREE: COMPROMISERS' in 1931 (B-LOB) vs. 1961 (LOB)

| | | | |
|---|---|---|---|
| crouch on the floor and stay | quite | still , and if the worst sho | quite |
| a moment she would come back | quite | confident to say that we sto | quite |
| , a mile away . " " That is | quite | right , " says she ; " but y | quite |
| closely viewed , he might be | quite | a plain young man . His eyes | quite |
| ake another in order to make | quite | sure that I had not been dec | quite |
| thers , seemed to have drunk | quite | enough , and they were talki | quite |
| le there , who were not only | rather | quiet and subdued but looked | rather |
| alf hoped that something not | quite | proper might happen at any m | quite |
| turned out afterwards to be | rather | expensive . She sipped her g | rather |
| ff' and I certainly found it | rather | exhilarating myself . The di | rather |
| . Perhaps I even swaggered , | to an extent | that I scarcely remember in | to an extent |
| back into her eyes and face | quite | steadily and with an express | quite |
| ge into the hotel . She felt | rather | queer as Harvey signed the v | rather |
| at this hotel before . It 's | rather | jolly , is n't it ? ' But Ja | rather |
| uppose you 're pleased . And | rather | superior about it , into the | rather |
| ng of their own later on . " | Quite | nice , " he said . " In fact | Quite |
| at would n't die , he knew , | rather | unhappily , that he was afra | rather |
| e Ways . It came to him with | rather | a shock , once he 'd kissed | rather |

## Example 2: Key concept OBLIGATION/NECESSITY in 20th C. BrE

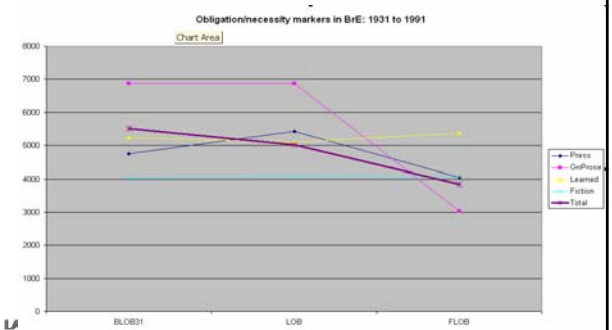
Background:

- Recent observations of significant shifts having occurred among expressions of obligation/necessity in the period 1961-1991 (Leech 2003, Smith 2003). E.g.
  - a decline of the central modals MUST and NEED
  - a spread of the semi-modals HAVE TO, NEED TO

**Questions**

- **Are these changes recent?**
- **How do these changes compare to the development of the semantic field of OBLIGATION/ NECESSITY *as a whole*?**

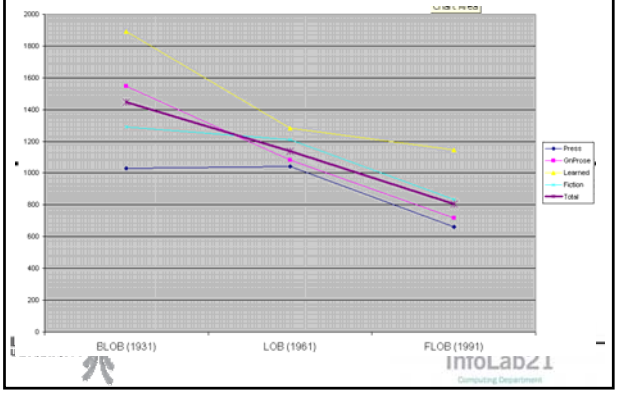
## Example 2: Key concept OBLIGATION/NECESSITY in 20th C. BrE




## 'OBLIGATION/NECESSITY' in 1931 Fiction

| | | | |
|---|---|---|---|
| ugh the sight reassured me , I | must | confess that I dreaded taking p | must |
| ut whether he was or was not I | should | not know till one of two things | should |
| e her , I began to fear that I | must | have passed her by : then the d | must |
| aid I , " I shall hear all you | have to | say : and you are to sit well b | have to |
| orward whatever befalls . If I | should | stop or be stopped , you are to | should |
| quite still , and if the worst | should | happen , you must swear that yo | should |
| the worst should happen , you | must | swear that you do not know me , | must |
| Now if I had used my wits , I | should | have known that Lelia , to whom | should |
| om roads meant nothing , would | have to | see about her if she was to tel | have to |
| I have plenty of time . But I | must | not take you further . As it is | must |
| home till long after dark . It | must | be six miles to Merring from wh | must |
| o see her . You think you will | have to | present me and that I am not fi | have to |
| tter , " said I. " And now you | must | go to your home . Soon after da | must |
| you do not come back ? " " You | must | go about your business , " said | must |
| id you good-bye to-day . " Why | should | you ? " said I. " If I do not c | should |
| lowered her eyes . " I do not | need | to see her , " she said . The l | need |
| o go home . ' ` I 'll go . You | need | n't , ' Jenny said , with decis | need |
| feet as slim . The beggar maid | must | have had a startling beauty in | must |
| ttention , but Jenny 's charms | needed | a close inspection . Kings did | needed |
| had given that thus her father | must | once have looked , and her own | must |

## MUST in BrE: 1931 to 1991



## Conclusion (1)

- Matrix extends keywords approach to key grammatical classes and key concepts using Log Likelihood to compare frequency profiles
- Key grammatical categories and semantic classes are used to group together lower frequency words and those words which would, by themselves, not be identified as key, and would otherwise be overlooked
- Comparison at the POS and semantic levels reduces the number of key categories that the researcher needs to examine

## Conclusion (2): Planned future developments

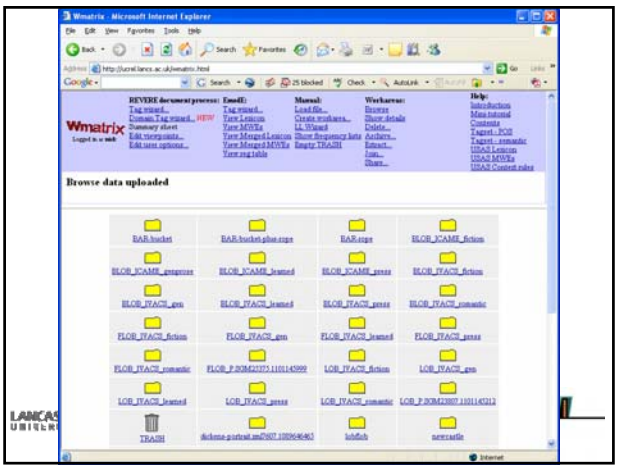
- Semantic tagging:
  - More accurate semantic disambiguation
  - A more refined semantic analysis for certain categories, e.g. deontic vs. epistemic use of MUST, HAVE TO etc.
- Semantic profiling
  - Incorporation of a Dispersion measure to filter 'key concepts' tables
  - Option to filter text according to markup types, e.g. quoted speech

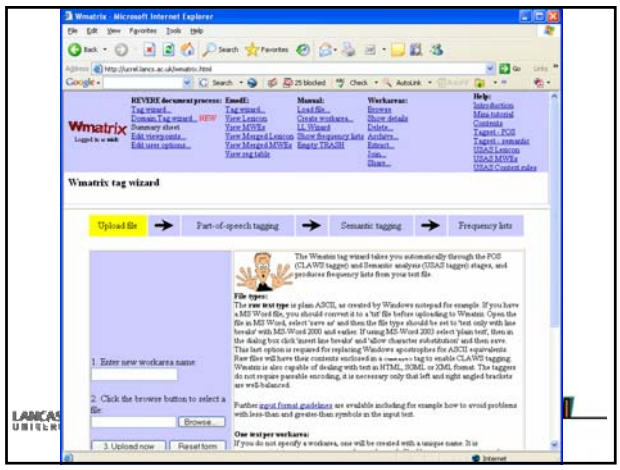
## Further info

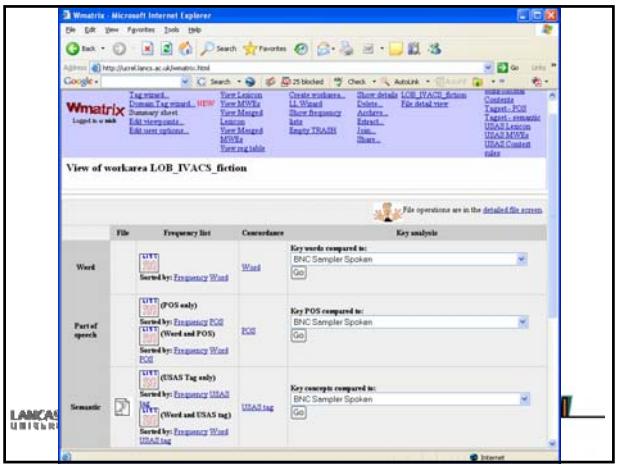
- http://ucrel.lancs.ac.uk/20thCenturyEnglish/
- Contacts: Paul Rayson (paul@comp.lancs.ac.uk), Nick Smith (nick@comp.lancs.ac.uk)
- Sponsored by The Leverhulme Trust (Grant number F/00 185/J), this project runs from August 2005 - July 2007.

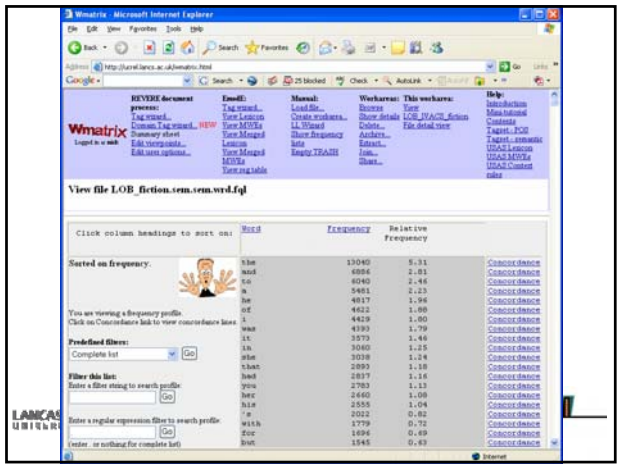
## References

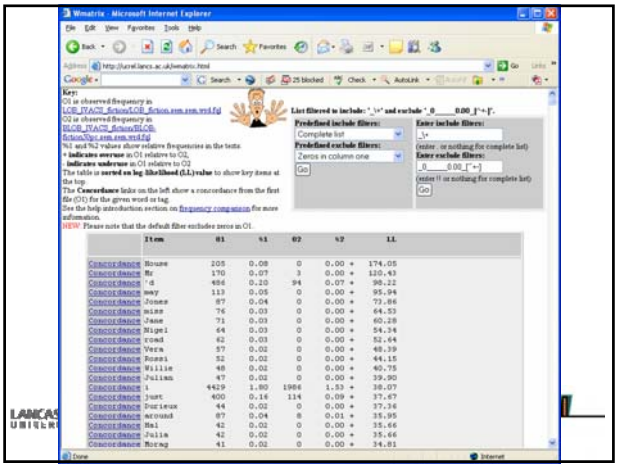

- Rayson, P. (2003). Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. *Ph.D. thesis*, Lancaster University.
- Rayson, P., Leech, G., and Hodges, M. (1997). Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British National Corpus. International Journal of Corpus Linguistics. Volume 2, number 1. pp 133 - 152. John Benjamins, Amsterdam/Philadelphia.
- Granger, S., and Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (ed.) Learner English on Computer. Longman, London and New York, pp. 119-131.
- Sawyer, P., Rayson, P. and Cosh, K. (2005) Shallow Knowledge as an Aid to Deep Understanding in Early Phase Requirements Engineering. IEEE Transactions on Software Engineering. Volume 31, number 11, November, 2005, pp. 969 - 981.

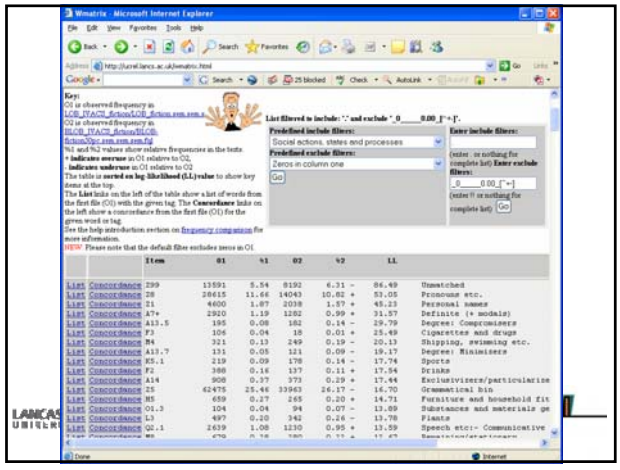



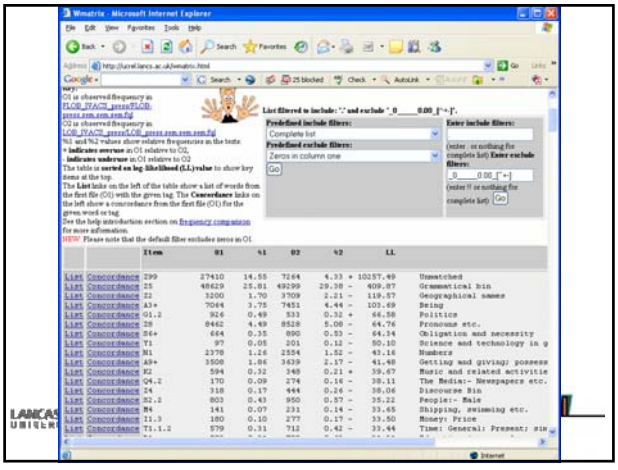



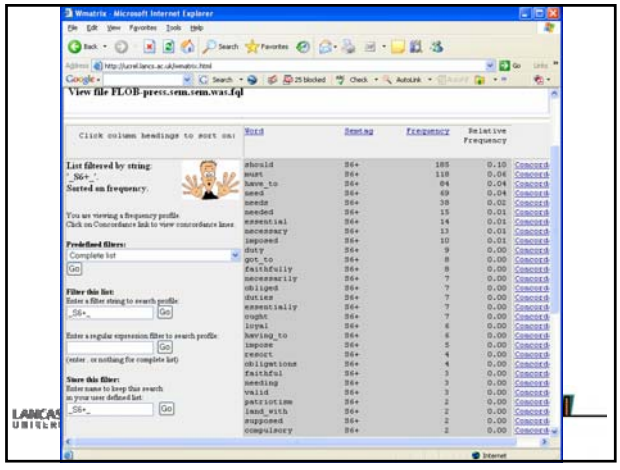

Wmatrix - Microsoft Internet Explorer
Wmatrix
View of workarea LOB_IVACS_fiction

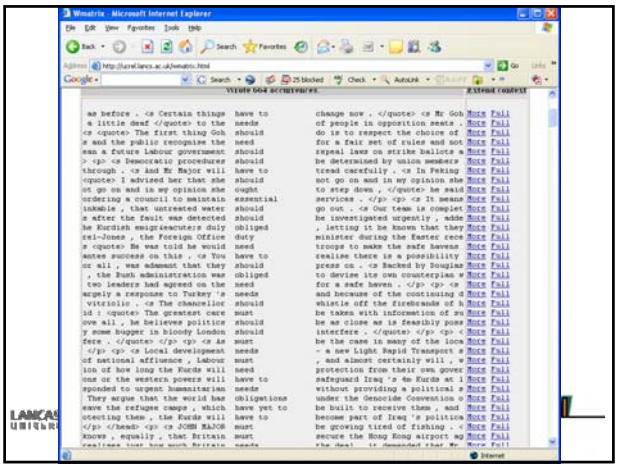

Wmatrix - Microsoft Internet Explorer
Wmatrix
View file LOB_fiction.sem.sem.wrd.fql


Wmatrix - Microsoft Internet Explorer


Wmatrix - Microsoft Internet Explorer


Wmatrix - Microsoft Internet Explorer


Wmatrix - Microsoft Internet Explorer
View file FLOB-press.sem.sem.wss.fql


Wmatrix - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://ucrel.lancs.ac.uk/wmatrix.html
Google Search 25 blocked Check AutoLink
Verbs 664 occurrences
Extend context
as before . <s Certain things have to change now . </quote> <s Mr Goh More Full
a little deaf </quote> to the needs of people in opposition seats . More Full
<s <quote> The first thing Goh should do is to respect the choice of More Full
s and the public recognise the need for a fair set of rules and not More Full
ean A future Labour government should repeal laws on strike ballots a More Full
>  <s Democratic procedures should be determined by union members More Full
through . <s And Mr Major will have to tread carefully . <s In Peking More Full
<quote> I advised her that she should not go on and in my opinion she More Full
ot go on and in my opinion she ought to step down , </quote> he said More Full
ordering a council to maintain essential services .   <s It means More Full
inkable , that untreated water should go out . <s Our team is complet More Full
s after the fault was detected should be investigated urgently , adde More Full
he Kurdish emigrés duly obliged , letting it be known that they More Full
rel-Jones , the Foreign Office duty minister during the Easter rece More Full
s <quote> He was told he would need troops to make the safe havens More Full
antee success on this . <s You have to realise there is a possibility More Full
or all , was adamant that they should press on . <s Backed by Douglas More Full
, the Bush administration was obliged to devise its own counterplan w More Full
two leaders had agreed on the need for a safe haven .   <s More Full
urgnly a response to Turkey 's needs and because of the continuing d More Full
vitriolic . <s The chancellor should whistle off the firebrands of h More Full
id : <quote> The greatest care must be taken with information of su More Full
ove all , he believes politic should be as close as is feasibly poss More Full
y some bugger in bloody London should interfere . </quote>   < More Full
fere . </quote>   <s As must be the case in many of the loca More Full
  <s Local development needs - a new Light Rapid Transport s More Full
of national affluence , Labour must , and almost certainly will , w More Full
ion of how long the Kurds will need protection from their own gover More Full
ons or the western powers will have to safeguard Iraq 's 4m Kurds at l More Full
sponded to urgent humanitarian needs without providing a political s More Full
They argue that the world has obligations under the Genocide Convention o More Full
eave the refugee camps , which have yet to be built to receive them , and More Full
otecting them , the Kurds will have to become part of Iraq 's politica More Full
 </head>  <s JOHN MAJOR must be growing tired of fishing . < More Full
knows , equally , that Britain must secure the Hong Kong airport ag More Full
realises just how much Britain needs the deal , it demanded that Mr More Full
Internet