Time's Pencil:

some signs of ageing in the writings of eminent authors

Richard Forsyth
Division of Clinical Neurosciences
University of Southampton
R.Forsyth@soton.ac.uk

Sonnet 16

Bvt wherefore do not you a mightier waie
Make warre vpon this bloudie tyrant time?
And fortifie your felfe in your decay
With meanes more bleffed than my barren rime?
Now ftand you on the top of happie houres,
And many maiden gardens yet vnset
With vertuous wifh would beare your liuing flowers,
Much liker than your painted counterfeit:
So fhould the lines of life that life repaire
Which this (Times pencil or my pupil pen)
Neither in inward worth nor outward faire
Can make you liue your felfe in eies of men.
  To giue away your felfe keeps your felfe ftill,
And you muft liue drawne by your owne fweet skill.

Outline

- Background
  - RF
  - "Cognitive Archaeology"
- Early results
- Future plans
- Discussion
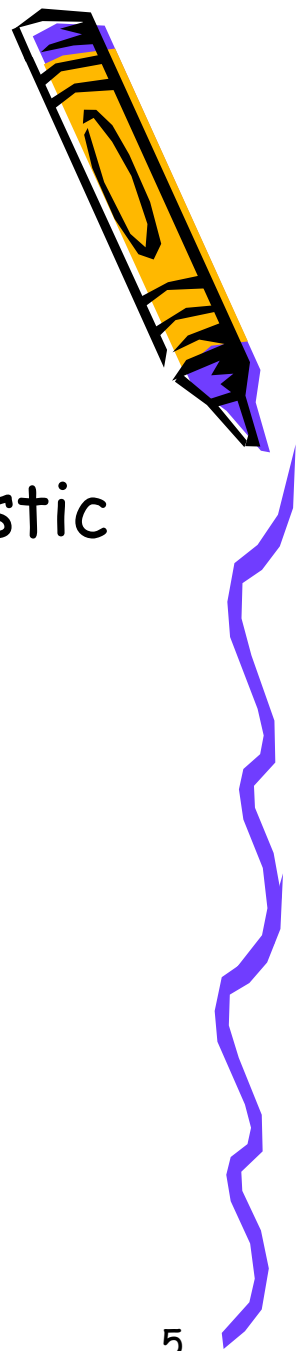 - feedback from today's audience
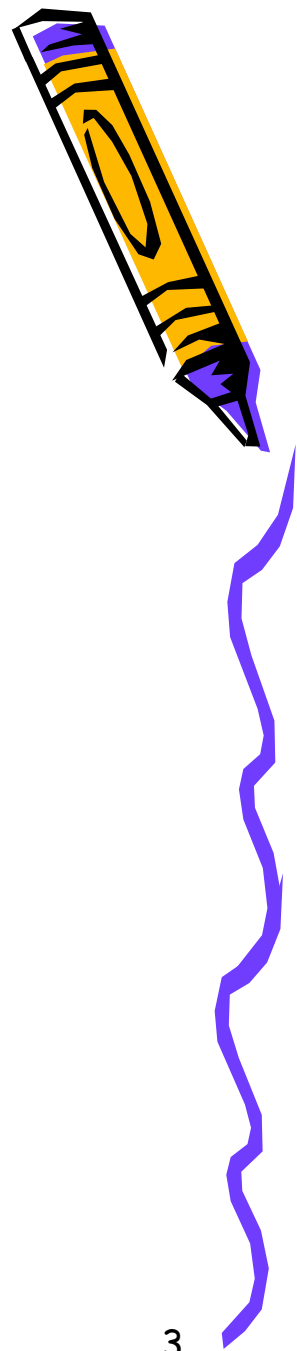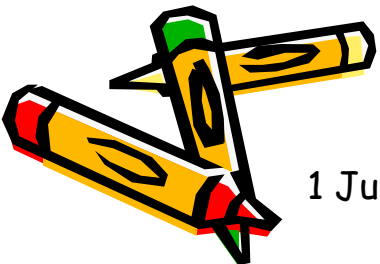
RF background

- Data Mining
  - Text Mining
    - Text categorization
      - Authorship attribution
    - Text evaluation
      - Stylochronometry
        » dating texts from stylistic evidence
        » Macro-stylochronometry (group)
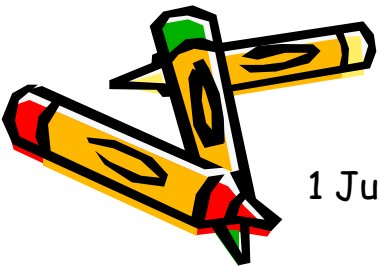   » Micro-stylochronometry (individual)
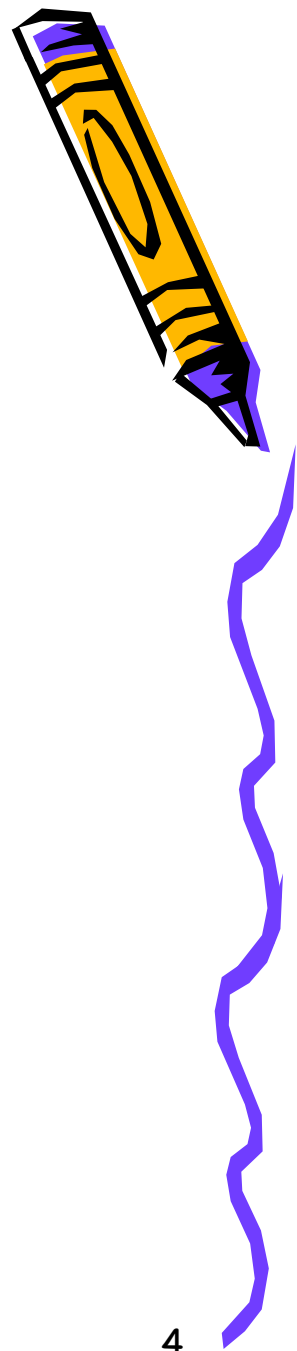
Text evaluation

- Finding a function/formula to predict a numeric attribute of a text from linguistic attributes of that text
  - Y = f(textual features)
- Stylochronometry:
  - Y = year (of composition)
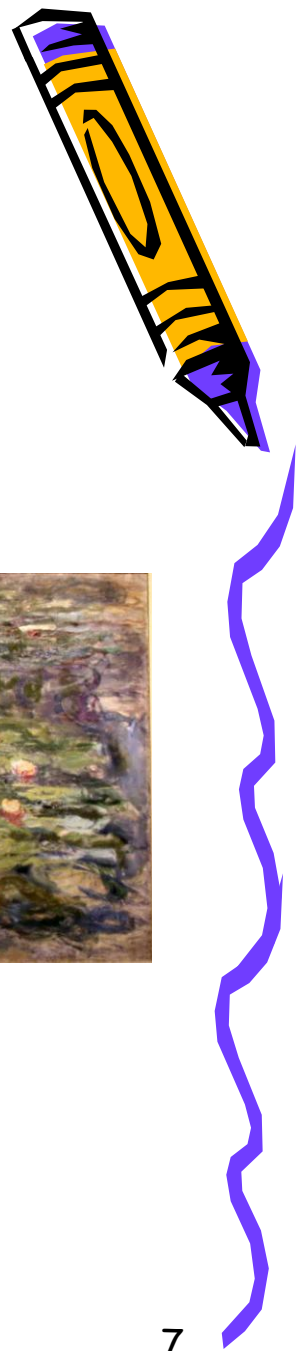- Medical linguistics:
  - Y = severity of condition

As the artist ages,

Time's Pencil: stylochronometry

does his art age too?

Time's Pencil: stylochronometry

Lily pond, 1926

Time's pencil writes lines on her face -
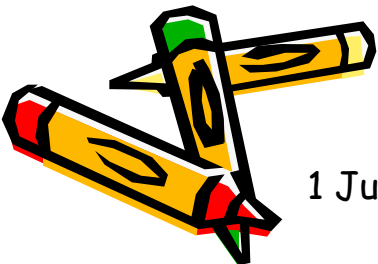
# does it affect lines written by her pencil (dictaphone)?
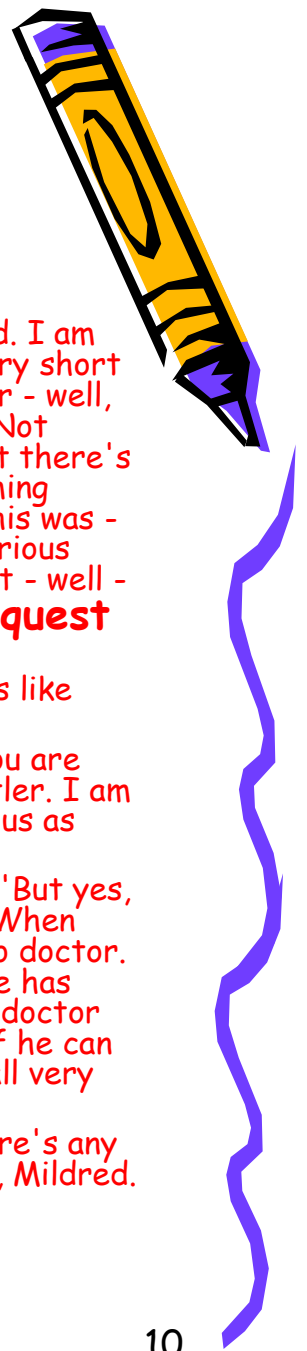
- We acquiesced and followed him out of the room. John strode on ahead and I took the opportunity of whispering to Poirot:

- "There will be an **inquest** then?"

- Poirot nodded absently. He seemed absorbed in thought; so much so that my curiosity was aroused.

- "What is it? You are not attending to what I say."

- "It is true, my friend. I am much worried."

- "Why?"

- "Because Mademoiselle Cynthia does not take sugar in her coffee."

- "What? You cannot be serious?"

- "But I am most serious. Ah, there is something there that I do not understand. My instinct was right."

- "What instinct?"

- "The instinct that led me to insist on examining those coffee-cups. Chut! no more now!"

- We followed John into his study, and he closed the door behind us.

- 'No, indeed,' said Henry. 'No, indeed. I am wondering really - yes, our time's very short you know - whether we hadn't better - well, give up this tour at this point here. Not continue with it. It seems to me that there's bound to be a bit of difficulty resuming things until we know definitely. If this was - well - I mean, if this should be so serious that it could prove fatal, there might - well - I mean there might have to be an **inquest** or something of that kind.'

- 'Oh Henry, don't say dreadful things like that!'

- 'I'm sure,' said Miss Cooke, 'that you are being a little too pessimistic, Mr Butler. I am sure that things couldn't be as serious as that.'

- In his foreign voice Mr Caspar said: 'But yes, they are serious. I hear yesterday. When Mrs Sandbourne talk on telephone to doctor. It is very, very serious. They say she has concussion bad - very bad. A special doctor he is coming to look at her and see if he can operate or if impossible. Yes - it is all very bad.'

- 'Oh dear,' said Miss Lumley. 'If there's any doubt, perhaps we ought to go home, Mildred. I must look up the trains, I think.'
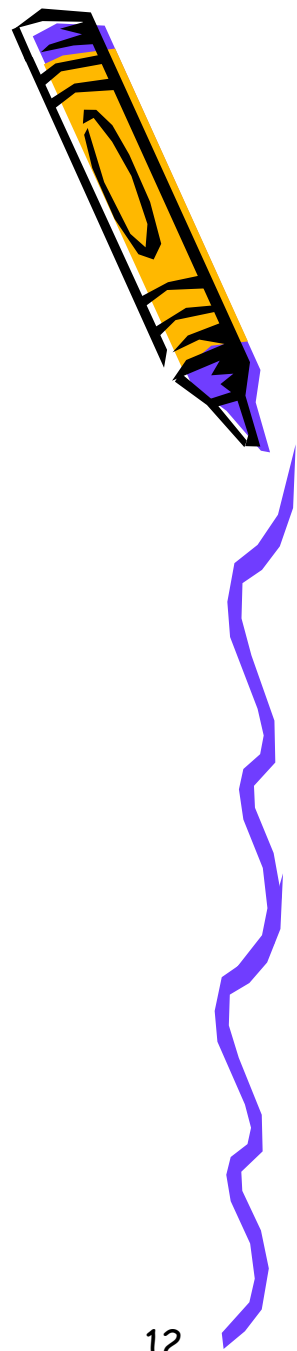
Project Background (PG)
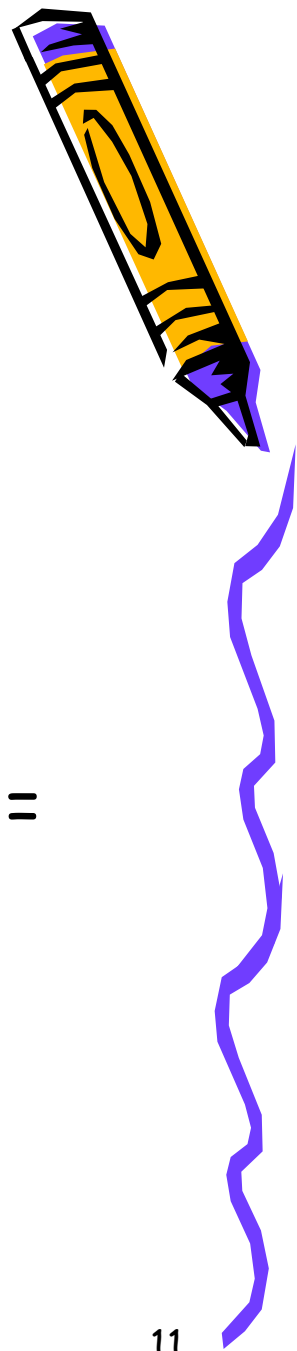
- Cognitive Archaeology
  - Diagnostic indicators from natural language
    - Focus on dementia
  - Verbal behaviour + linguistic analysis = early diagnosis! (?)
- -> Text mining with a medical slant
 - (initial emphasis on feature-finding)

Language certainly changes, for many reasons

- Fashion / Social change
  - I am, sir, your most humble & obedient servant -> Cheers
  - Bombay -> Mumbai (1996); Burma -> Myanmar (1989)
  - Negro, thou; chav, parenting; email, gay, mouse, web
  - "I'm like 'It's like, not expensive'."
- New technologies / discoveries
  - Bloggin' all over the world
  - satnav
- New media
  - Quill, fountain pen, pencil, ballpoint, typewriter, dictaphone, laptop
  - C u l8r
- Personal development
  - New skills, new friends
  - Relocation
  - Traumatic events
  - Memory deficit
  - Etc.

Text sets for testing
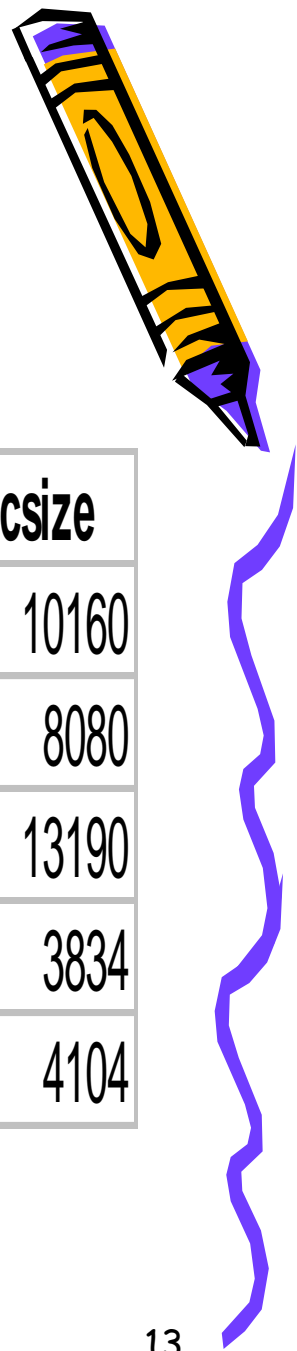
Name	Content	Docs	Words	Mean size	Vocsize
AC	Chapters by Agatha Christie	52	141093	2713.33	10160
Augs	US presidential inagural speeches	39	102012	2615.69	8080
IM	Chapters* by Iris Murdoch	52	213492	4105.62	13190
WY	Poems by WB Yeats	89	19919	223.81	3834
Xmas	Christmas broadcasts by QE 2	57	37797	663.11	4104

An example from politics

- My fellow citizens:

- I stand here today humbled by the task before us, grateful for the trust you have bestowed, mindful of the sacrifices borne by our ancestors. I thank President Bush for his service to our nation, as well as the generosity and co-operation he has shown throughout this transition.

- Forty-four Americans have now taken the presidential oath.

- [....]

- Thank you. God bless you. And God bless the United States of America.

Another example, from Politics

politics

in our time *the* destiny of *man* presents *its* meaning in political terms .

-- thomas mann

how *can* i *,* *that* girl standing there *,*

my attention fix

on roman *or* on russian

or on spanish politics ?

yet here's *a* travelled *man that* knows

what he talks about *,*

and there's *a* politician

that has *both* read *and thought* ,

and maybe *what* they say *is* true

of war *and* war's alarms *,*

but o *that* i were *young* again

and held her in my arms !

…. X .. O . O ……… O . X O … X ….. O .. O …… O . O O . O … X X . O . O . O . X O X X . O .. O … X .. X O . O . O .. X . X ……
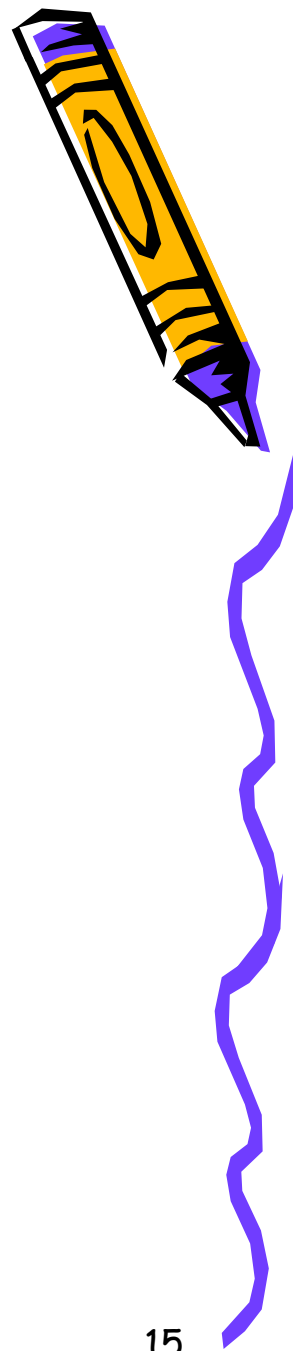
US inaugurals: negative & positive terms

•	which	−0.77976	−1	• old	0.61984	20
•	the	−0.71457	−2	• but	0.62955	19
•	mind	−0.67449	−3	• americans	0.63057	18
•	state	−0.65445	−4	• come	0.63077	17
•	of	−0.65344	−5	• is	0.63219	16
•	may	−0.65101	−6	• need	0.63947	15
•	wise	−0.64696	−7	• way	0.64008	14
•	duties	−0.63907	−8	• our	0.64069	13
•	constitution	−0.63745	−9	• america	0.64291	12
•	public	−0.61964	−10	• know	0.65607	11
•	by	−0.61903	−11	• are	0.67996	10
•	powers	−0.61599	−12	• work	0.69008	9
•	blessings	−0.60121	−13	• do	0.69899	8
•	objects	−0.60061	−14	• :	0.70061	7
•	union	−0.59575	−15	• new	0.70425	6
•	to	−0.59413	−16	• world	0.71154	5
•	under	−0.56822	−17	• today	0.75364	4
•	from	−0.55344	−18	• we	0.81154	3
•	general	−0.55304	−19	• god	0.82389	2
•	executive	−0.54150	−20	• .	0.89899	1

In "god" we trust

US Inaugurals, 1789-2009: %rate of 'god'.
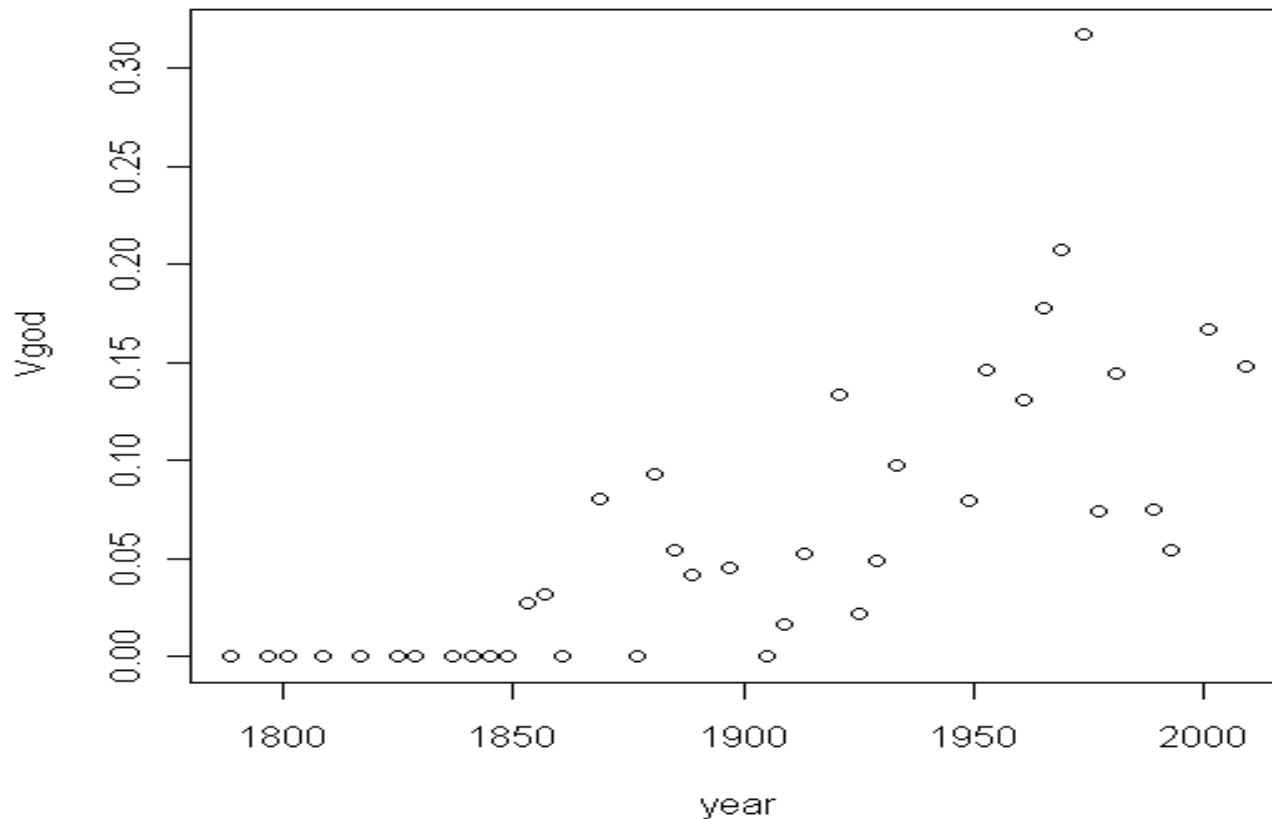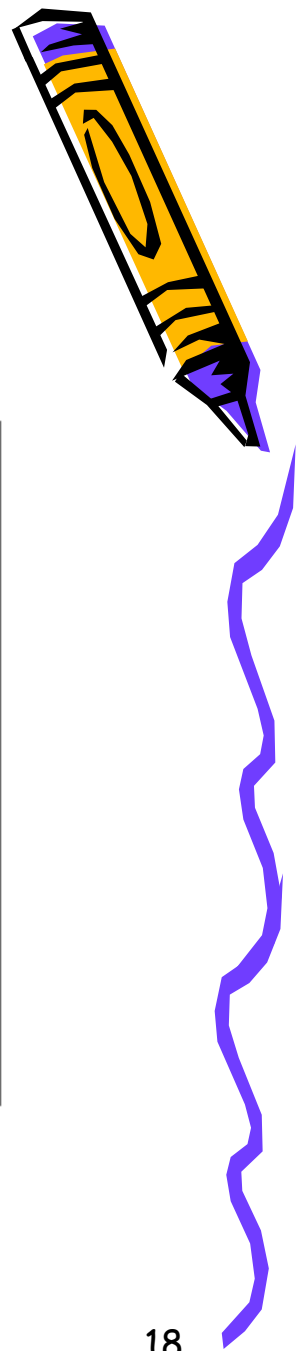
Every which way but up

US Inaugurals, 1789-2009: %rate of 'which'.
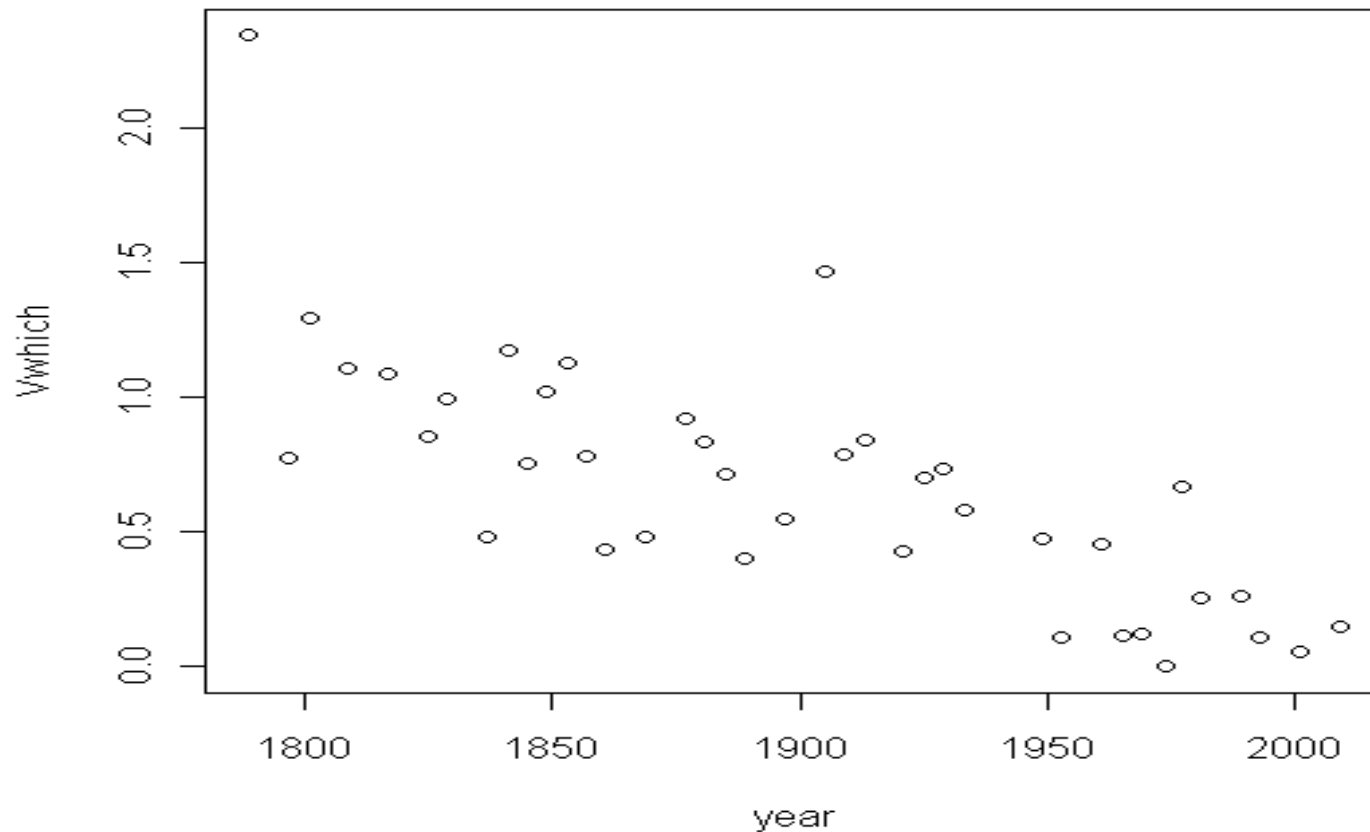
Periods over a 220-year period

US Inaugurals: will they ever stop?
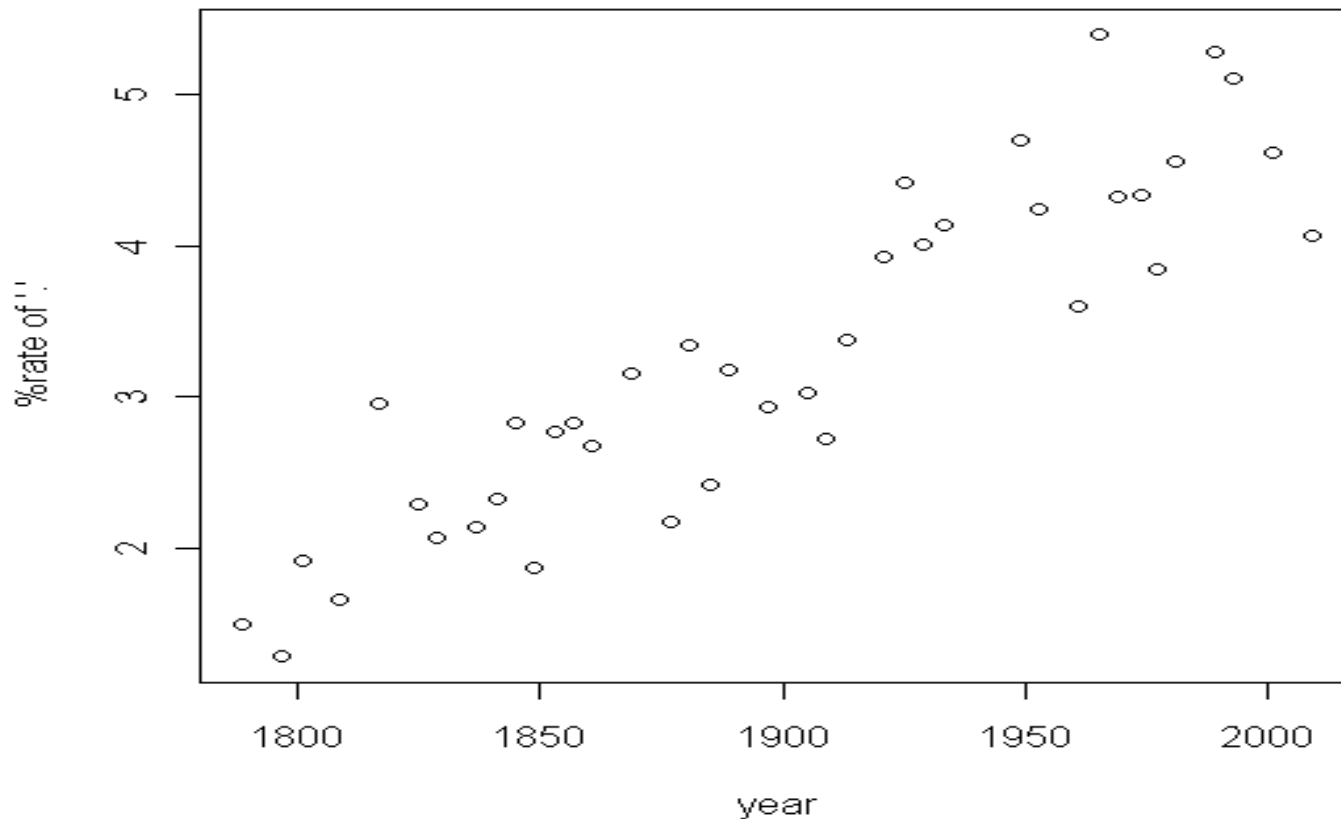
The meaning of Christmas?

QE2 Xmas broadcasts : 'jesus' or 'christ'.
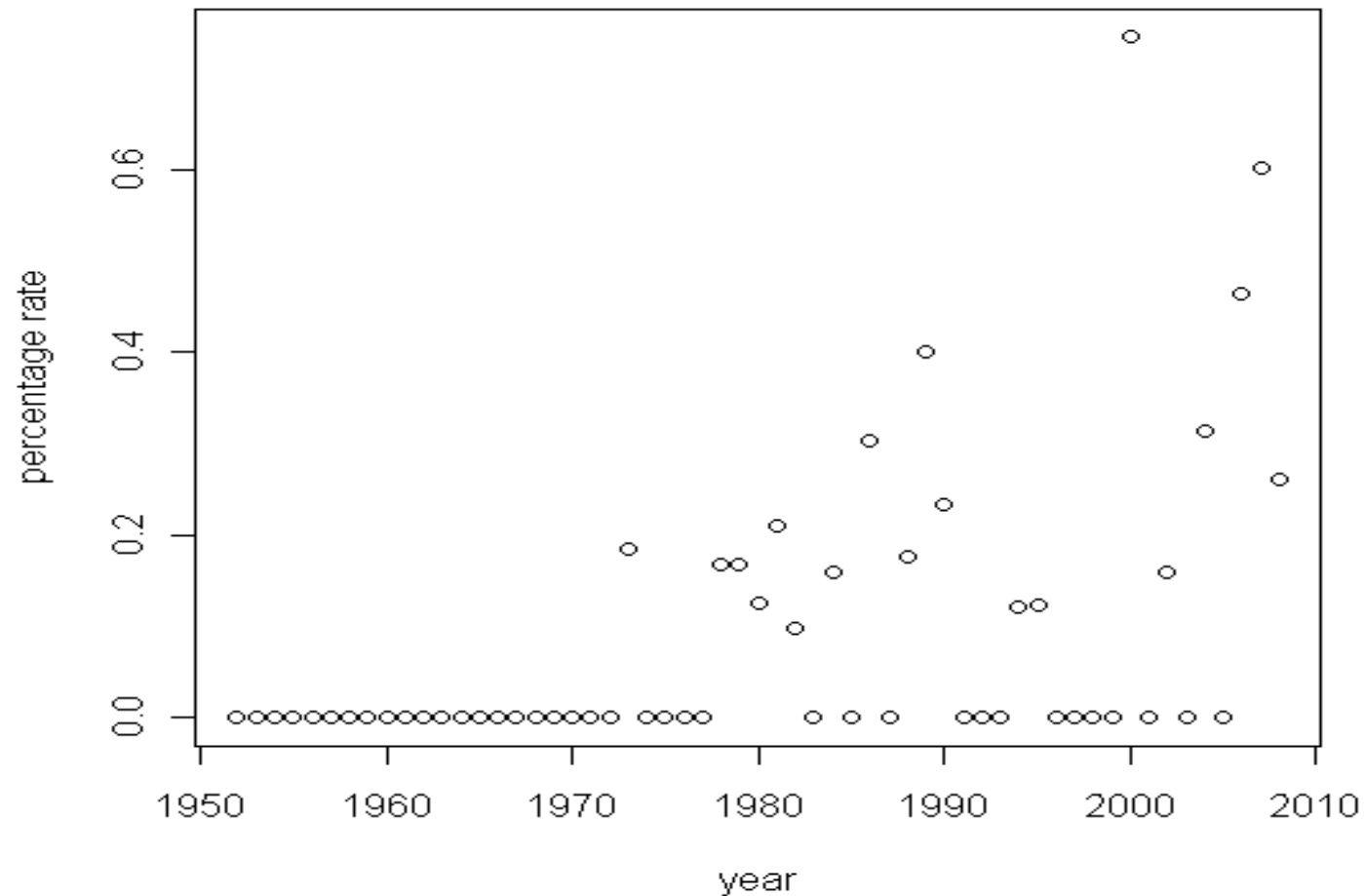
Post-imperial decline?

QE2 Xmas broadcasts : 'commonwealth'
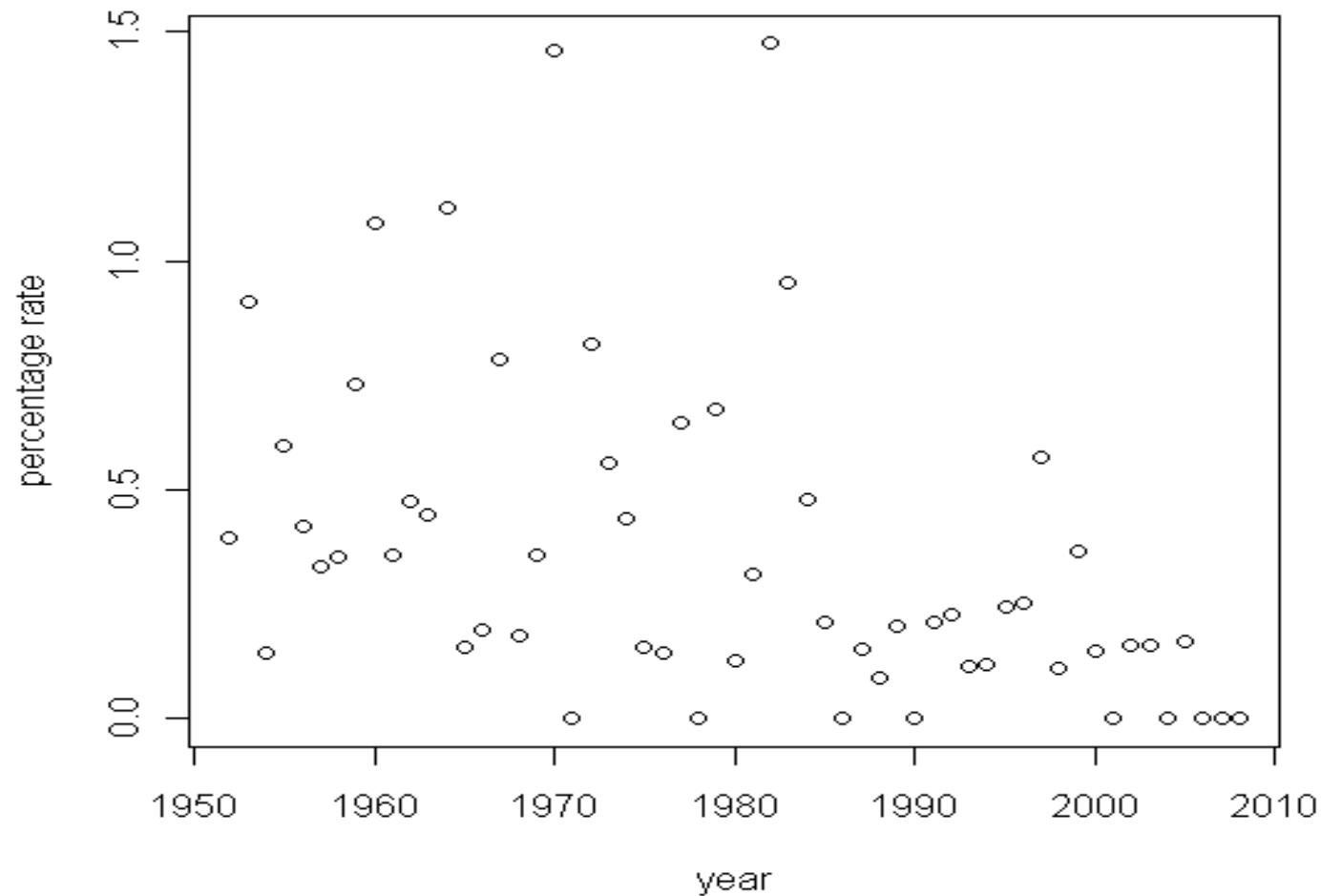
It's 'all' going downhill

QE2 Xmas broadcasts: 'all'.

"My Philip and I …"

QE2 Xmas broadcasts : o=husband, x=philip.
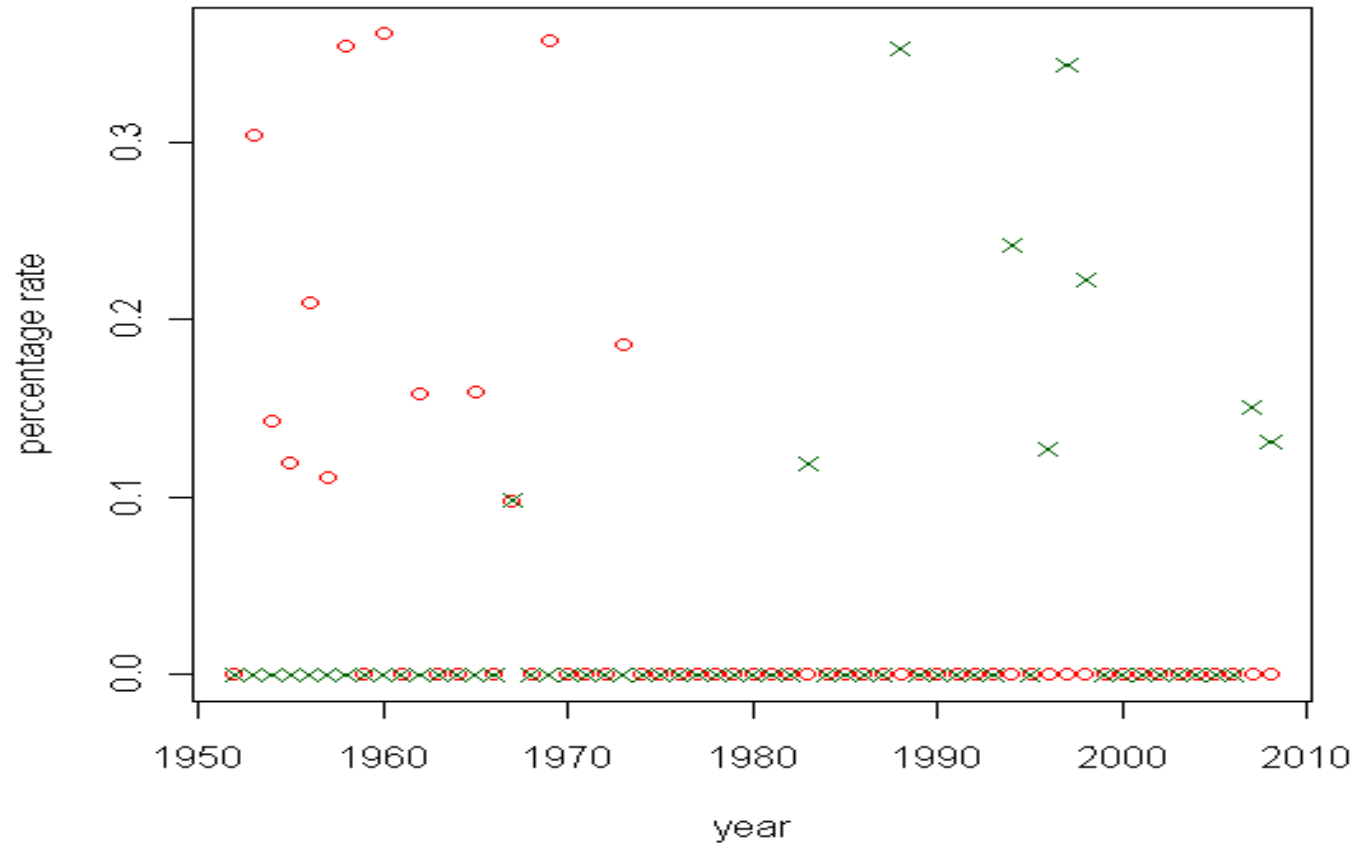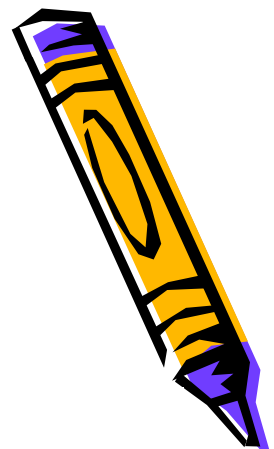
Agatha & Iris: 'or'

Agatha chapters: rate of 'or'.
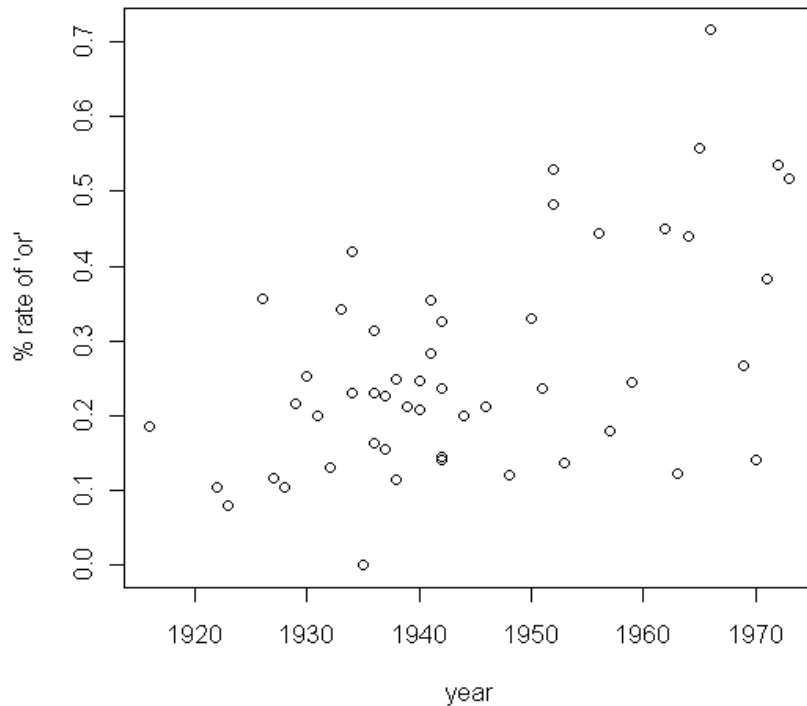
Iris chapters: rate of 'or'.
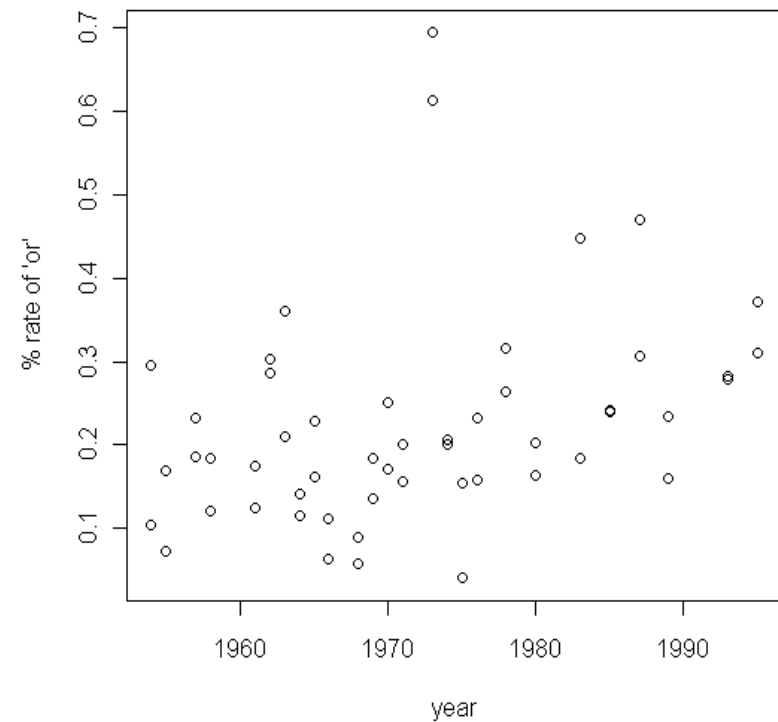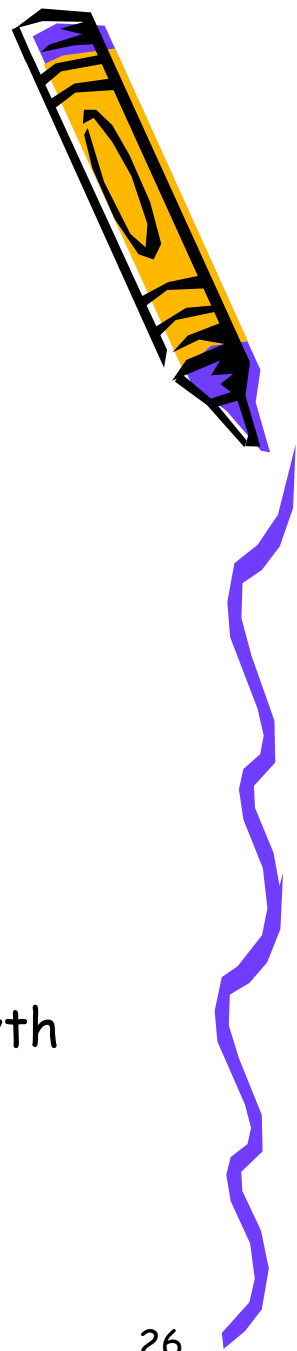
Agatha & Iris: dissolution of the ego
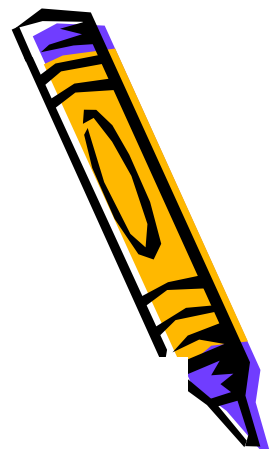
Agatha: 'me' + 'my'.
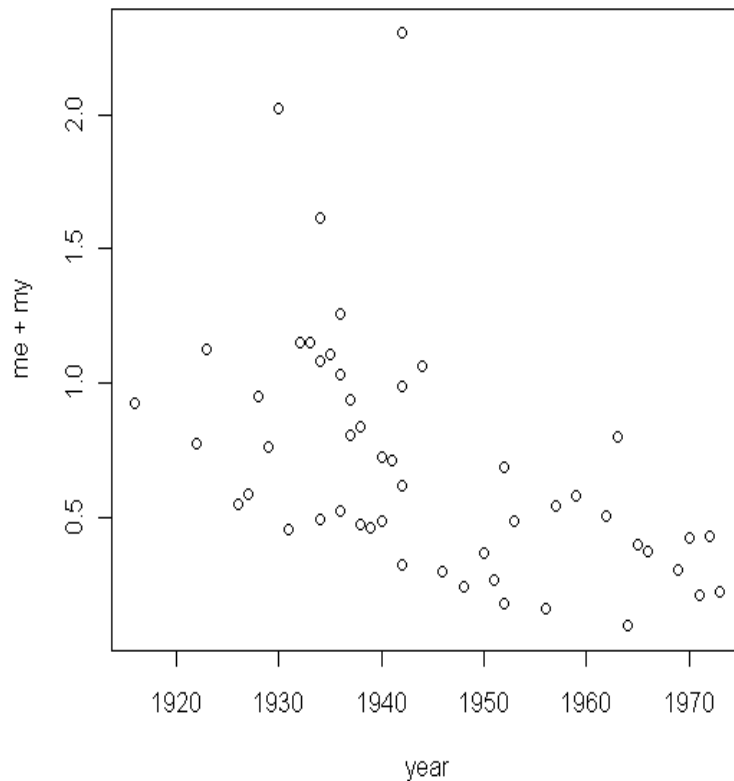
Iris: 'me' + 'my'.
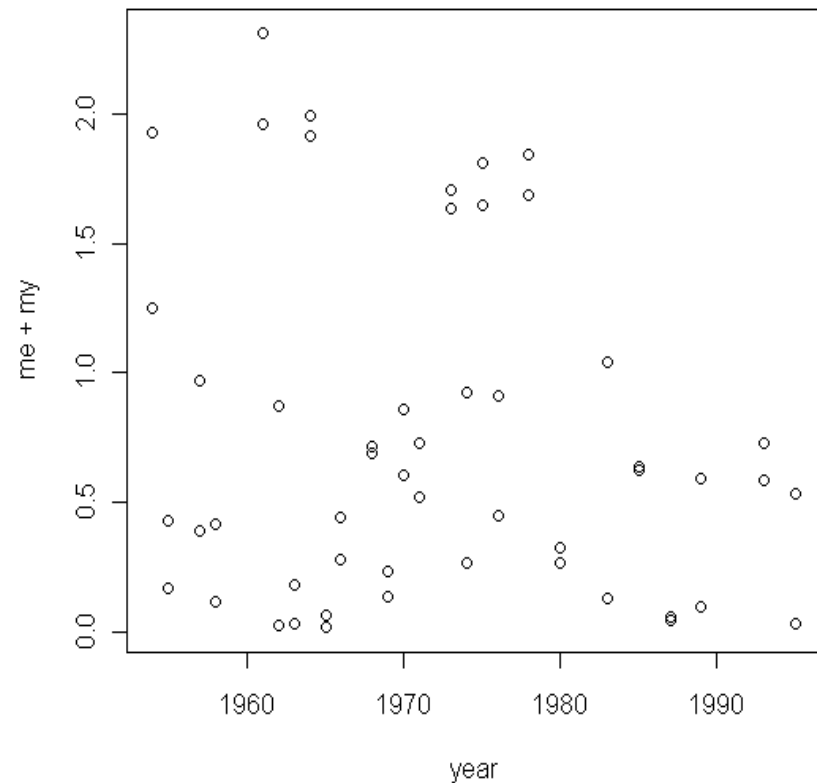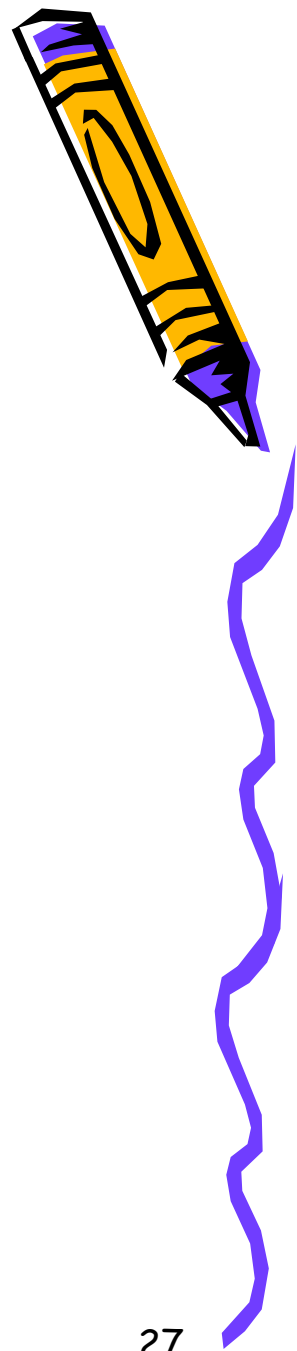
Time's Pencil: stylochronometry

Text-scoring procedure: basic idea

- Emphasize feature finding:
  - home-brew tokenizer, allows punctuation
    - still single-term based
  - select list of positive and negative markers
    - Association with y-value (e.g. year of composition)
- Use simple predictive formula:
  - Docscore = (P - N) / (P + N)
    - $P = \sum f(p_i) + 1$
    - $N = \sum f(n_i) + 1$
- Rationale:
  - Dawes (1979); Gigerenzer & Goldstein (1996); Forsyth (1999); Tetlock et al. (2007)
  - avoids "double" cross-validation
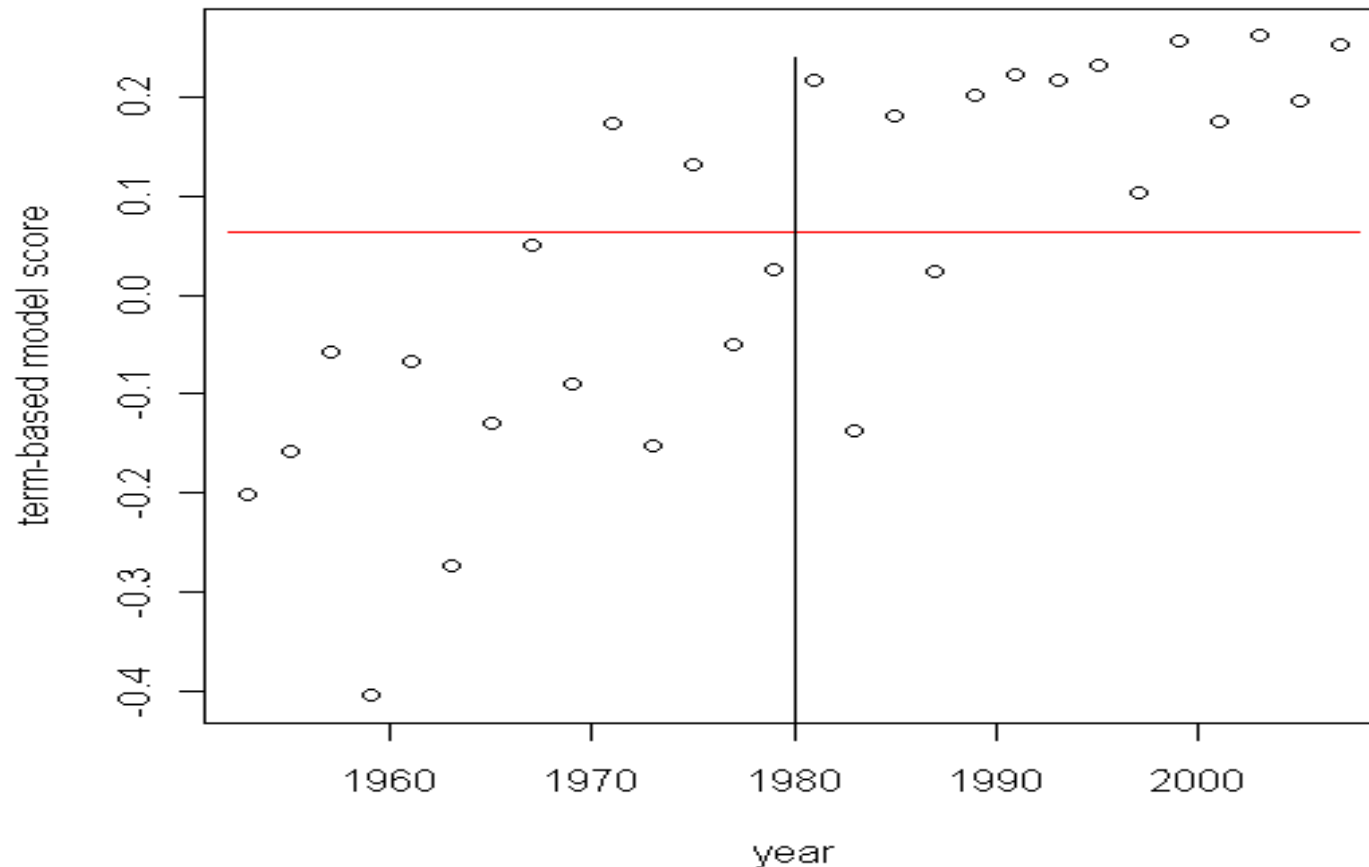
Agatha & Iris: something in common?

Agatha predicted by Iris model (r=0.54)
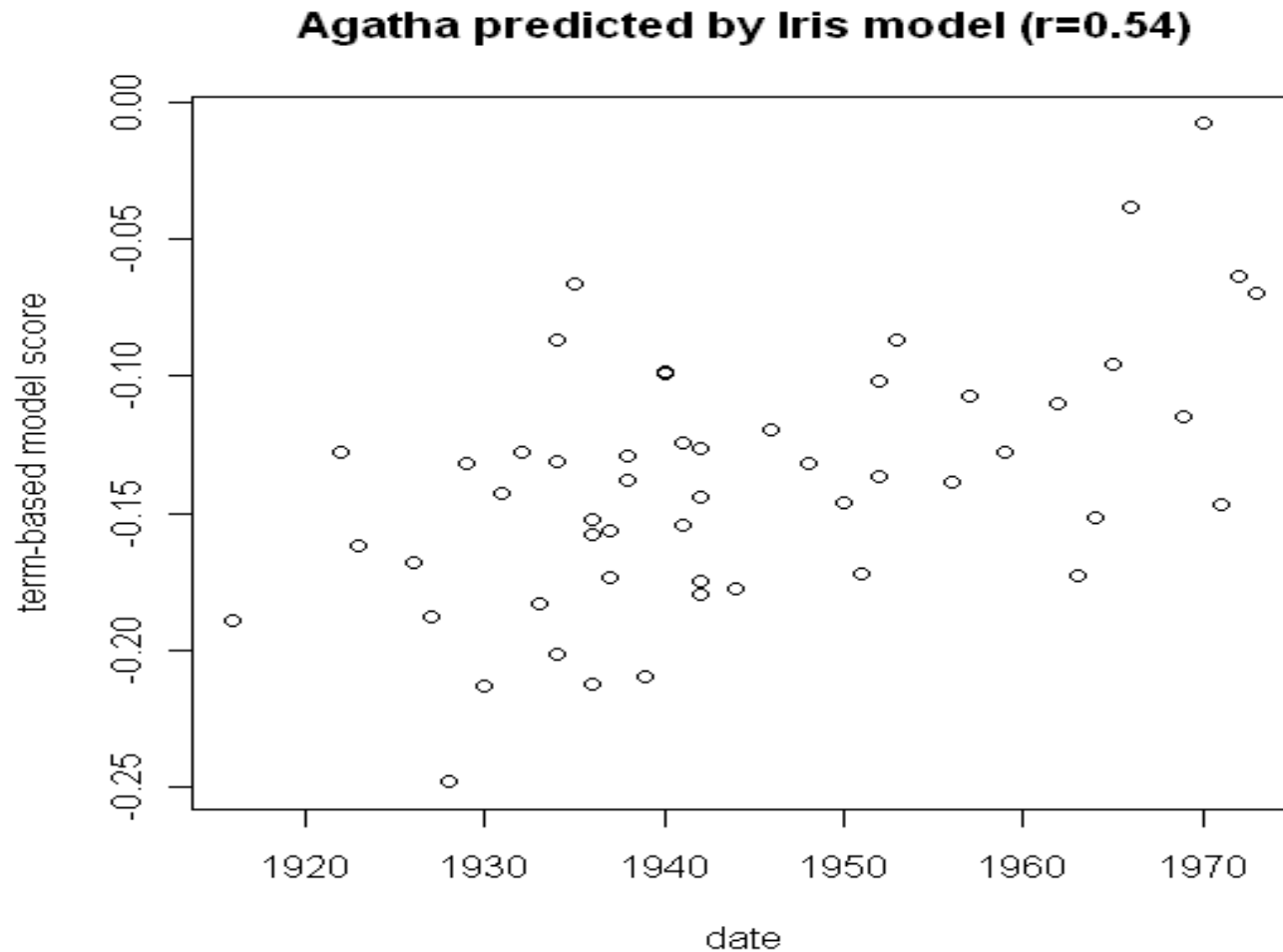
Xmas broadcasts: odd predicted by even

QE2 Xmas broadcasts (even->odd, keymode 2, r=0.8).

Dates of Yeats: holdout sample

WY13, holdout sample: r=0.87
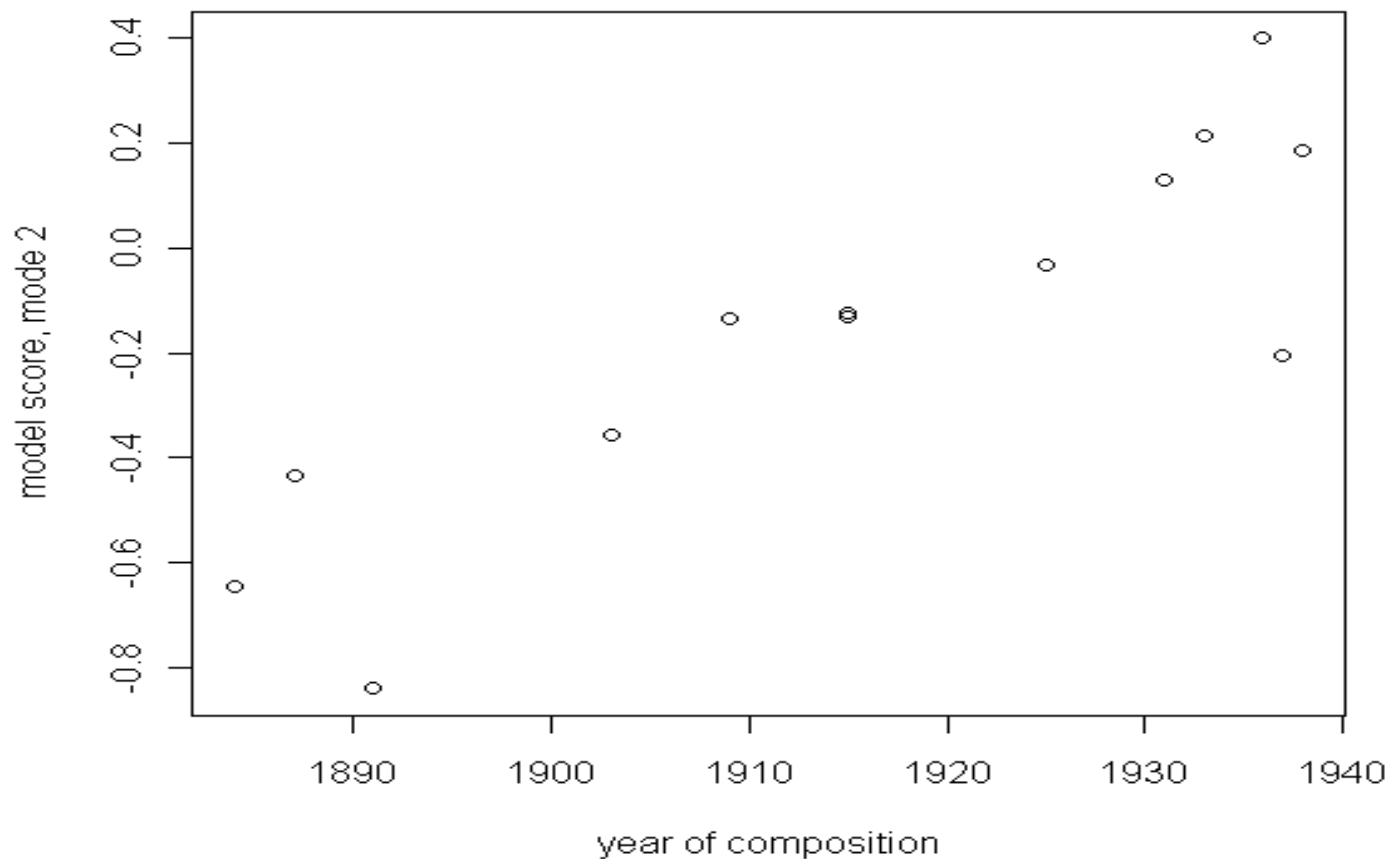
Agatha Christie: decline in diversity

AC : a diversity score that works (r=-0.52)
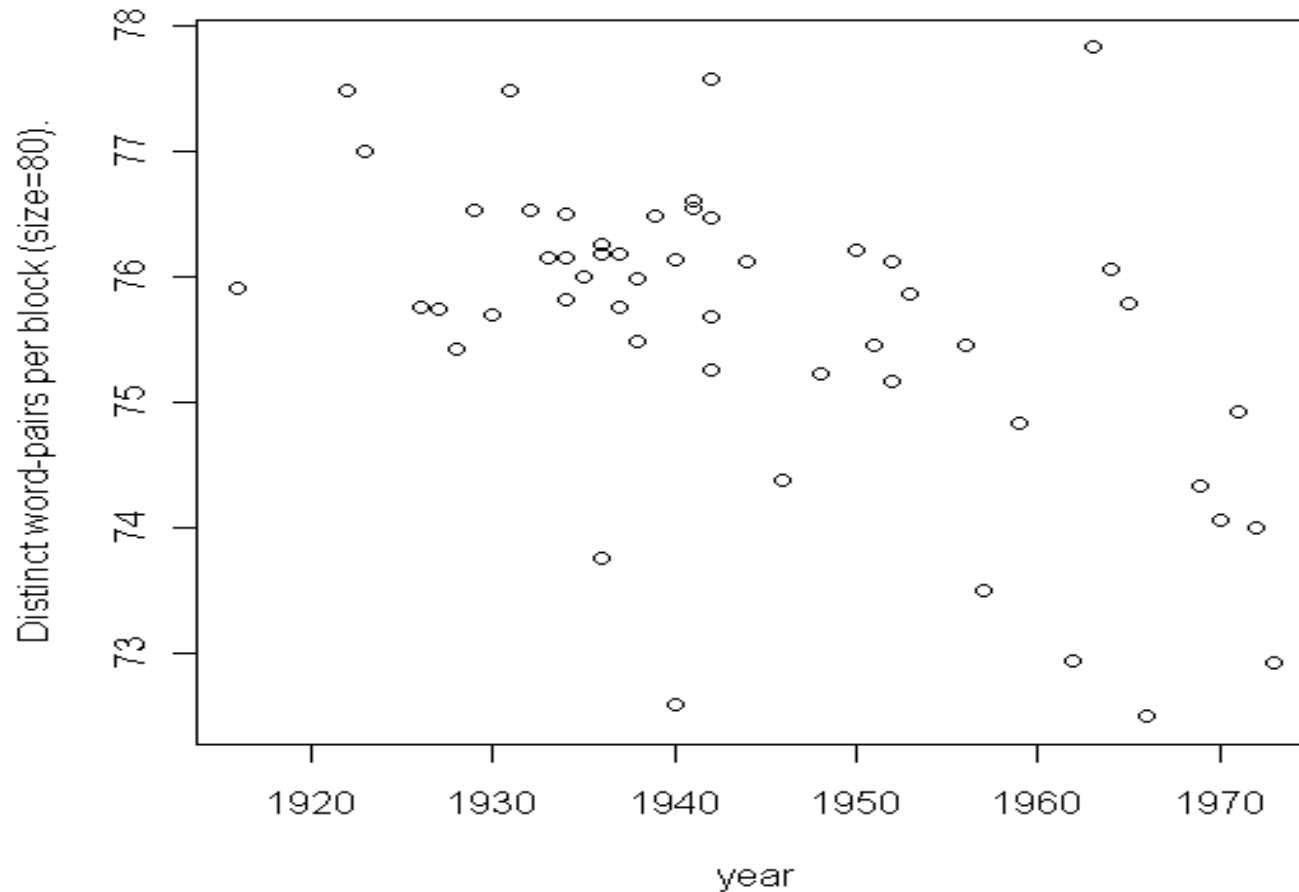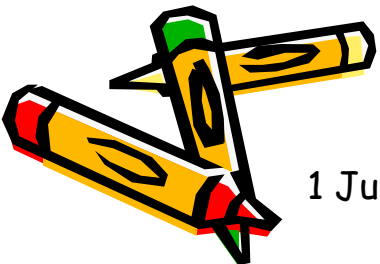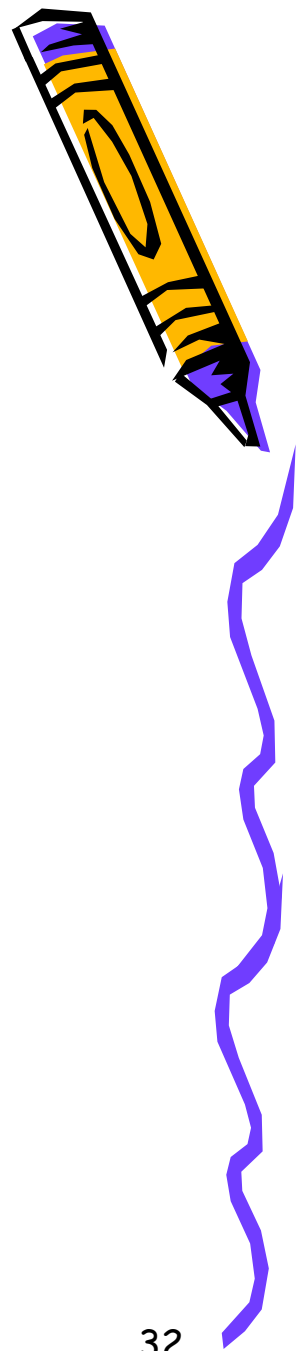
In conclusion

- No conclusions yet!
- Hints:
  - Many small effects, no single feature
  - Individual authors can be dated (approximately) but not all with the same indicators
  - "blurry" depictions (Monet) => "vague" descriptions (Agatha, Iris) ?

A long way to go ….
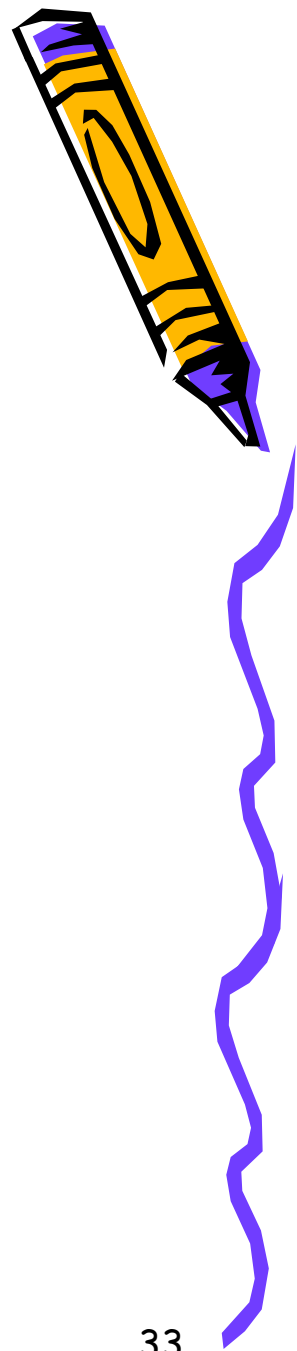
- Still to do:
    - Corpus compilation
        - Optima data
            - Cookie theft transcription
        - Appeal for donors!
    - Corpus "cleaning"
    - Exploring a range of indicators
        - Term rates
        - Diversity indices
        - Collocations & colligations
        - Etc.
    - Integrating the results

That's all folks!

- Thank you for your attention
- ☺