

Phraseology and diversity in student writing

Phil Durrant

p.l.durrant@exeter.ac.uk



Motivation

Variation as a key 'problem' in English for
Academic Purposes



Motivation

Can corpora help our understanding?



More specifically

What are disciplines?

How do they vary?

How do they relate to each other?

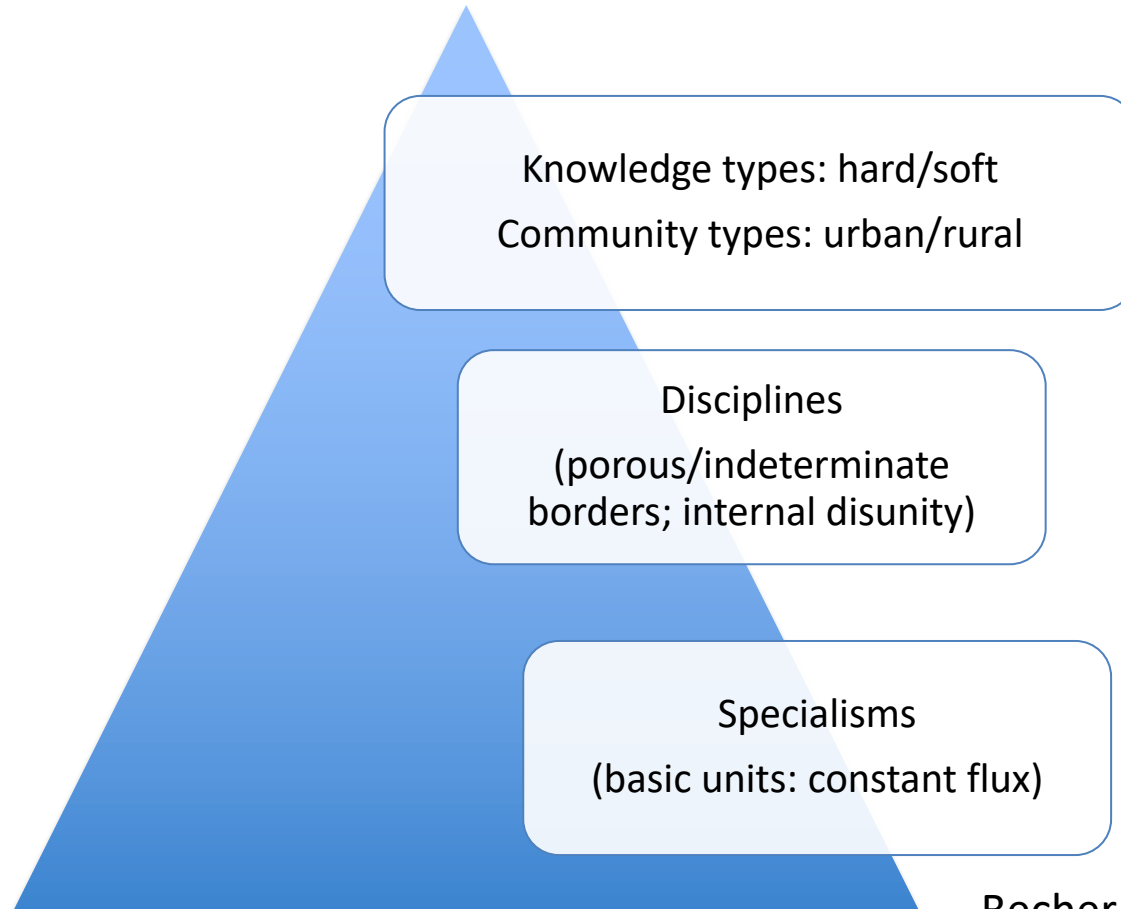


The construct of 'discipline'

- Are disciplines coherent units?



The construct of 'discipline'



Becher & Trowler 2001



The construct of 'discipline'

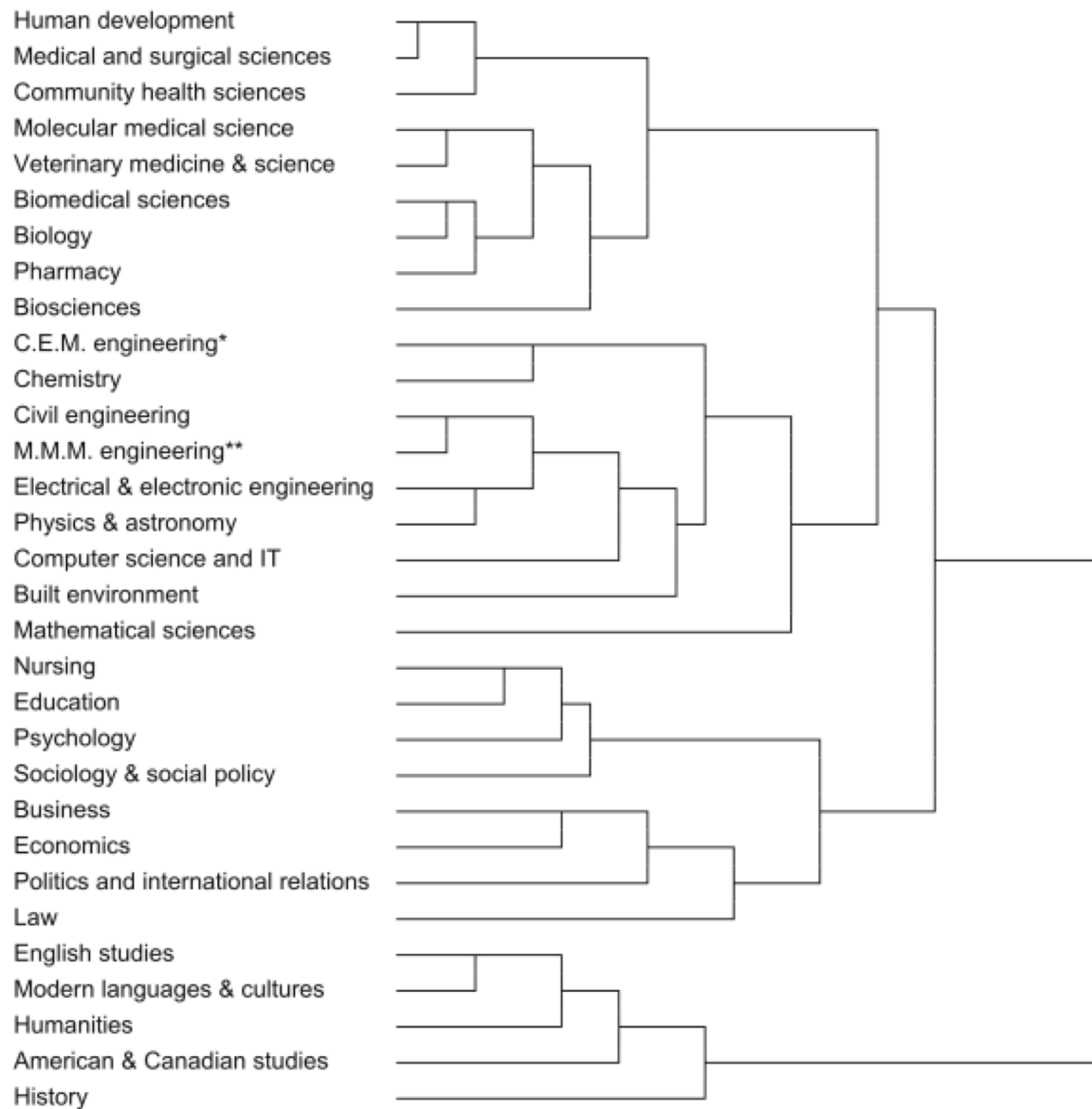
- Are disciplines coherent units?
- 'Negotiable', 'permeable', 'transient'? (Brew, 2008)
- Contextually-dependent:
 - Research; teaching; administration
 - Textual/linguistic classification??



Disciplines in corpus linguistics

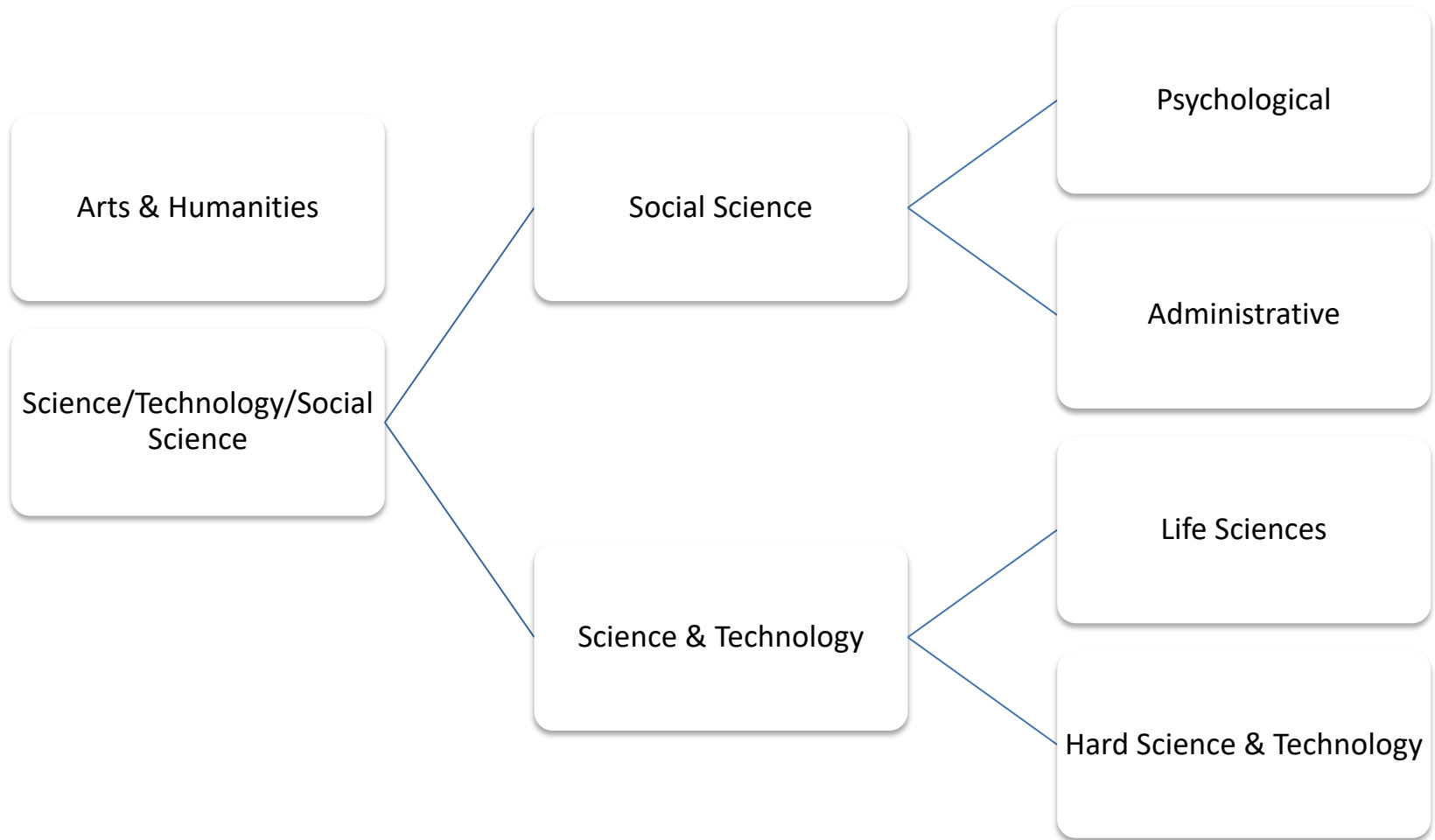
Source	Disciplinary Categories						
Coxhead (2000)	Arts		Commerce	Law		Science	
Biber (2006)	Humanities	Business	Education	Social Science	Natural Science	Engineering	
Hyland (2008)	Applied Linguistics		Business Studies		Microbiology	Electrical Engineering	
Simpson-Vlach & Ellis (2010)	Humanities & Arts		Social Sciences		Natural Sciences/Medicine		Technology & Engineering
Gardner & Davies (2014)	Humanities	History		Education	Social Science		Philosophy, religion and psychology
	Law & Political Science		Business & Finance		Medicine & Health		Science & Technology





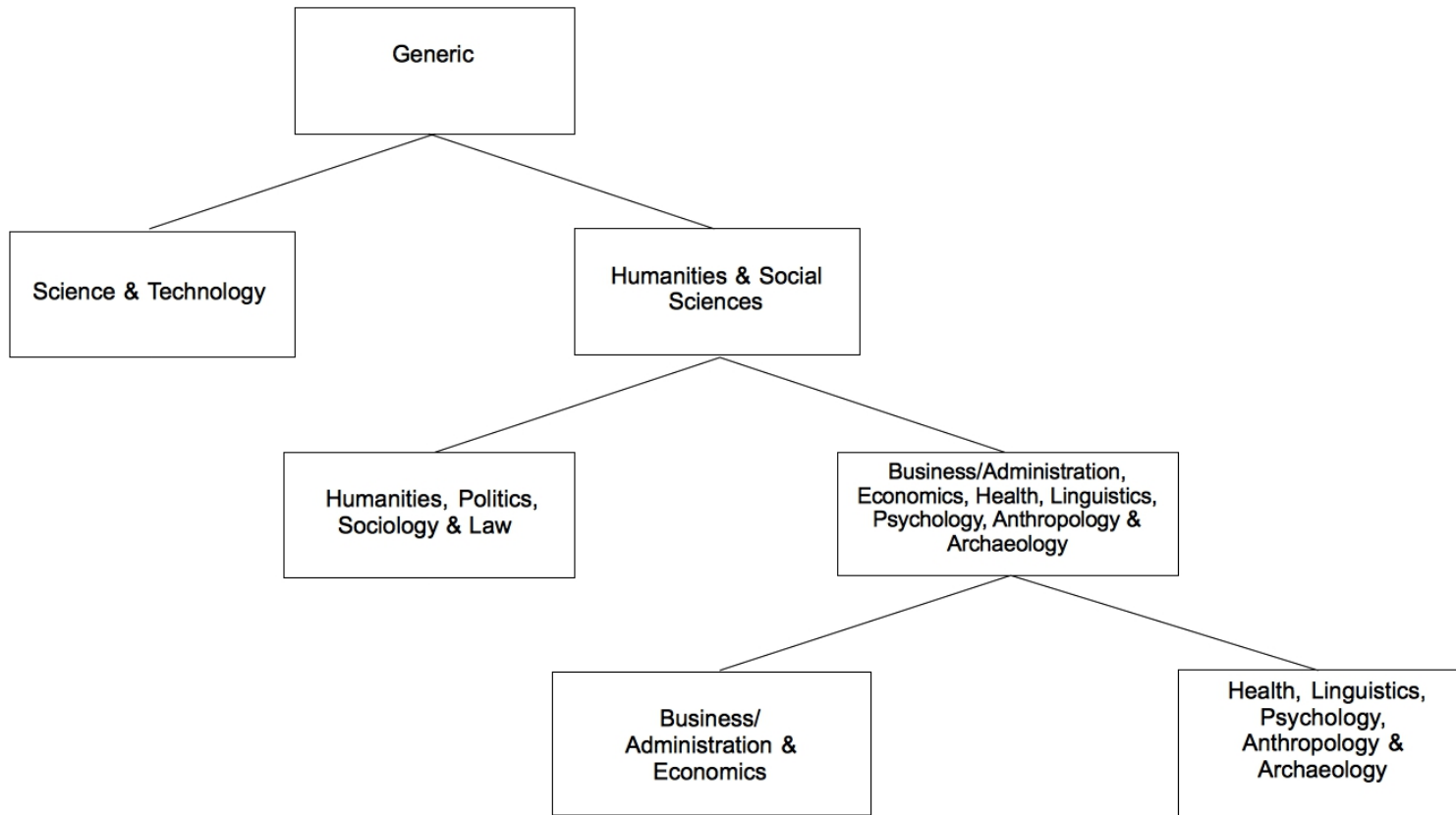
Durrant, 2009





Durrant, 2009





Durrant, 2014



- Based on word forms only – high degree of ambiguity
- Takes given disciplinary labels for granted



4-grams and diversity

- Automatic analysis of large numbers of texts
- Qualitatively interpretable
- Sensitive to contextual variation



Corpus

- British Academic Written English Corpus (Alsop & Nesi 2009)
- Divided into texts written by each author
- Authors with < 3 texts/6,000 words excluded
- Authors with multiple disciplines excluded
- 285 authors; 24 disciplines; 1,558 texts; 1,478 titles



Analysis

- Writers assigned IDs: e.g. *agriculture_6001*
- Each writer's text broken down into 4-grams...



Poland is predominantly an agricultural country, it is also a transition economy, World Trade Organization (WTO) member and a new European Union (EU) Member State. From all those reasons agri-food policies seem to be a fascinating topic for discussion. More than 15 years ago (in 1989), Poland regained its democracy and since then has been undergoing deep changes in all sectors of economy.



Poland is predominantly an agricultural country, it is also a transition economy, World Trade Organization (WTO) member and a new European Union (EU) Member State. From all those reasons agri-food policies seem to be a fascinating topic for discussion. More than 15 years ago (in 1989), Poland regained its democracy and since then has been undergoing deep changes in all sectors of economy.



Poland is predominantly an agricultural country, it is also a transition economy, World Trade Organization (WTO) member and a new European Union (EU) Member State. From all those reasons agri-food policies seem to be a fascinating topic for discussion. More than 15 years ago (in 1989), Poland regained its democracy and since then has been undergoing deep changes in all sectors of economy.



Analysis

- Writers assigned IDs: e.g. *agriculture_6001*
- Each writer's text broken down into 4-grams, e.g.:
 - poland is predominantly an
 - is predominantly an agricultural
 - predominantly an agricultural country
- Gives 285 lists of 4-grams (one per writer)



Analysis

- Find overlaps between each writer:
 - *Overlap = total shared 4-grams/total 4-grams*
- Create matrix of overlaps:

◇	A	B	C	D	E
1	id	agriculture_6001.txt	agriculture_6015.txt	agriculture_6031.txt	agriculture_6034.txt
2	agriculture_6001.txt	1	0.001482321	0.002013648	0.003024453
3	agriculture_6015.txt	0.001482321	1	0.001717892	0.001718565
4	agriculture_6031.txt	0.002013648	0.001717892	1	0.003095975
5	agriculture_6034.txt	0.003024453	0.001718565	0.003095975	1

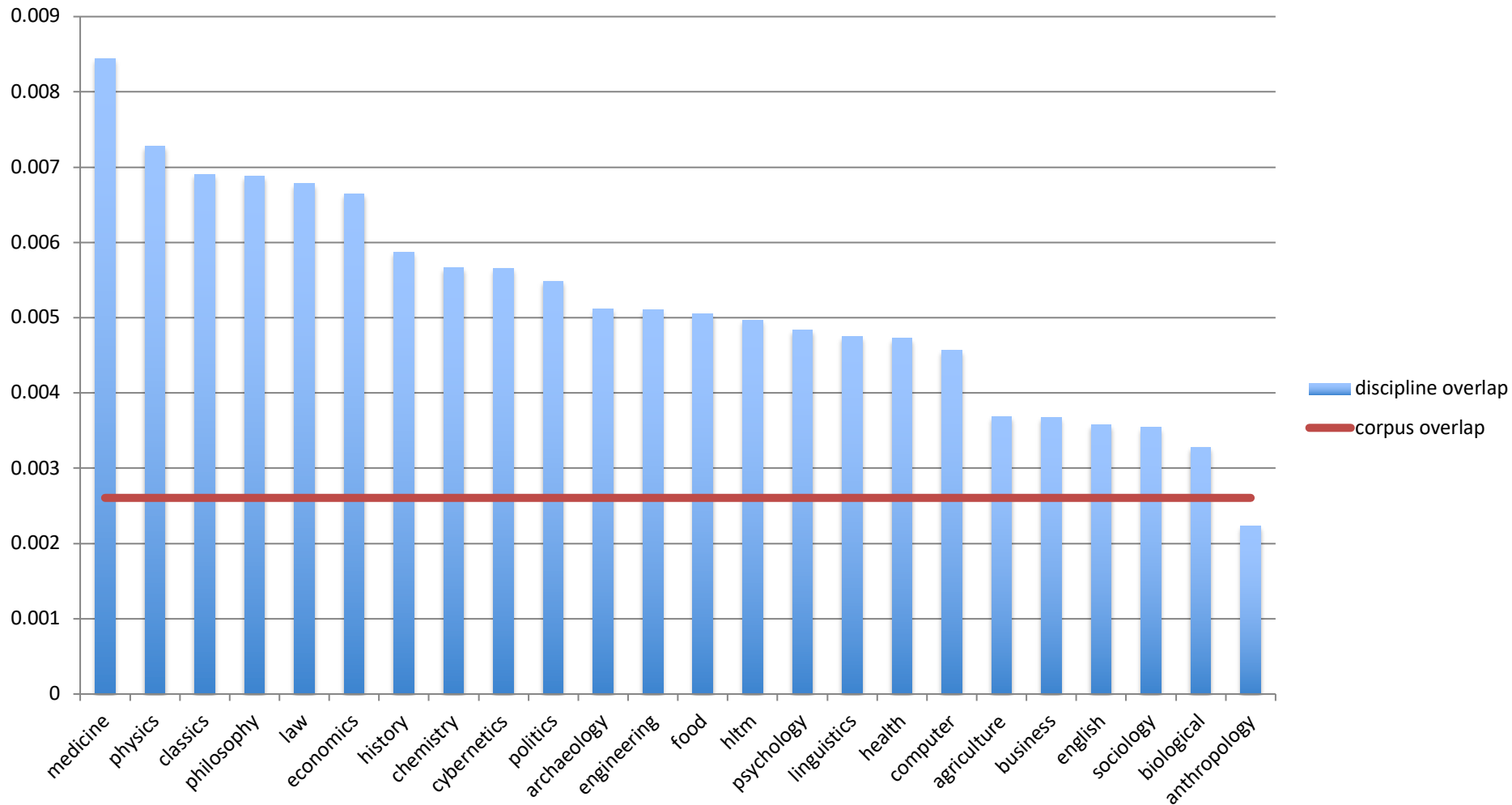


4-grams, lexical bundles and gradability

- No frequency threshold for overlap analysis
- All repeated 4-grams included
- More frequent bundles weigh more heavily
- ‘Lexical bundles’ analyzed at later stage



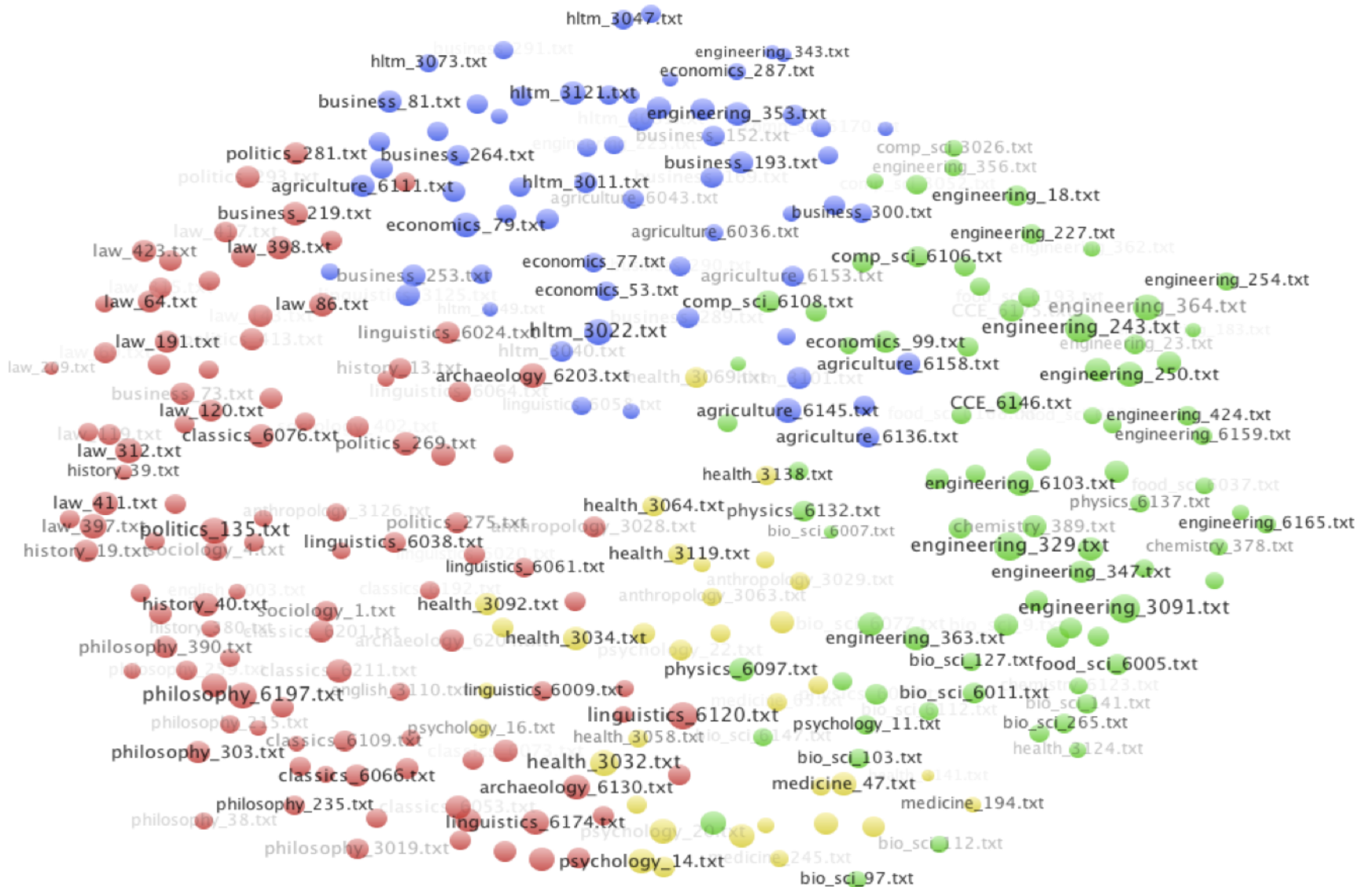
How homogenous are disciplines?

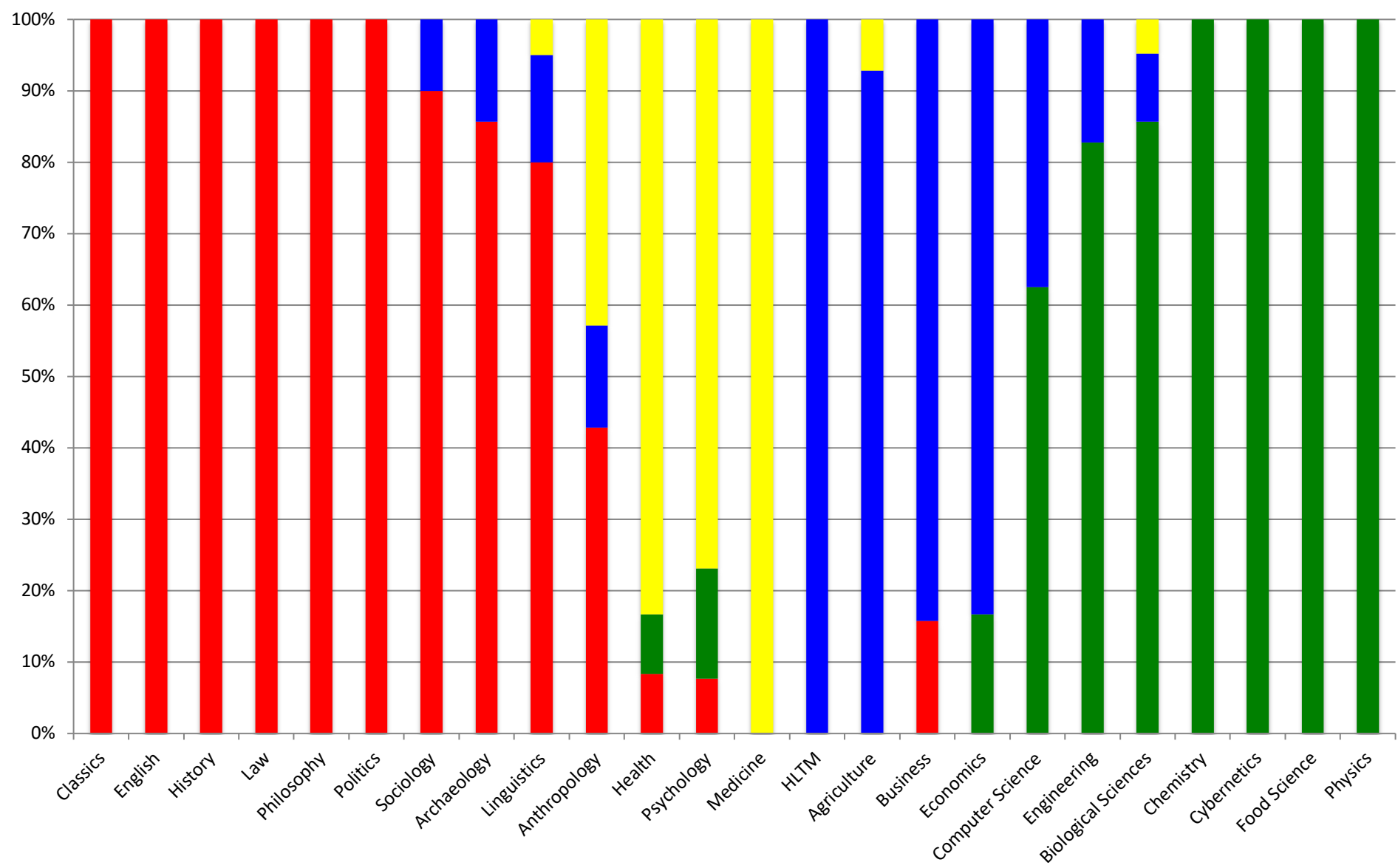


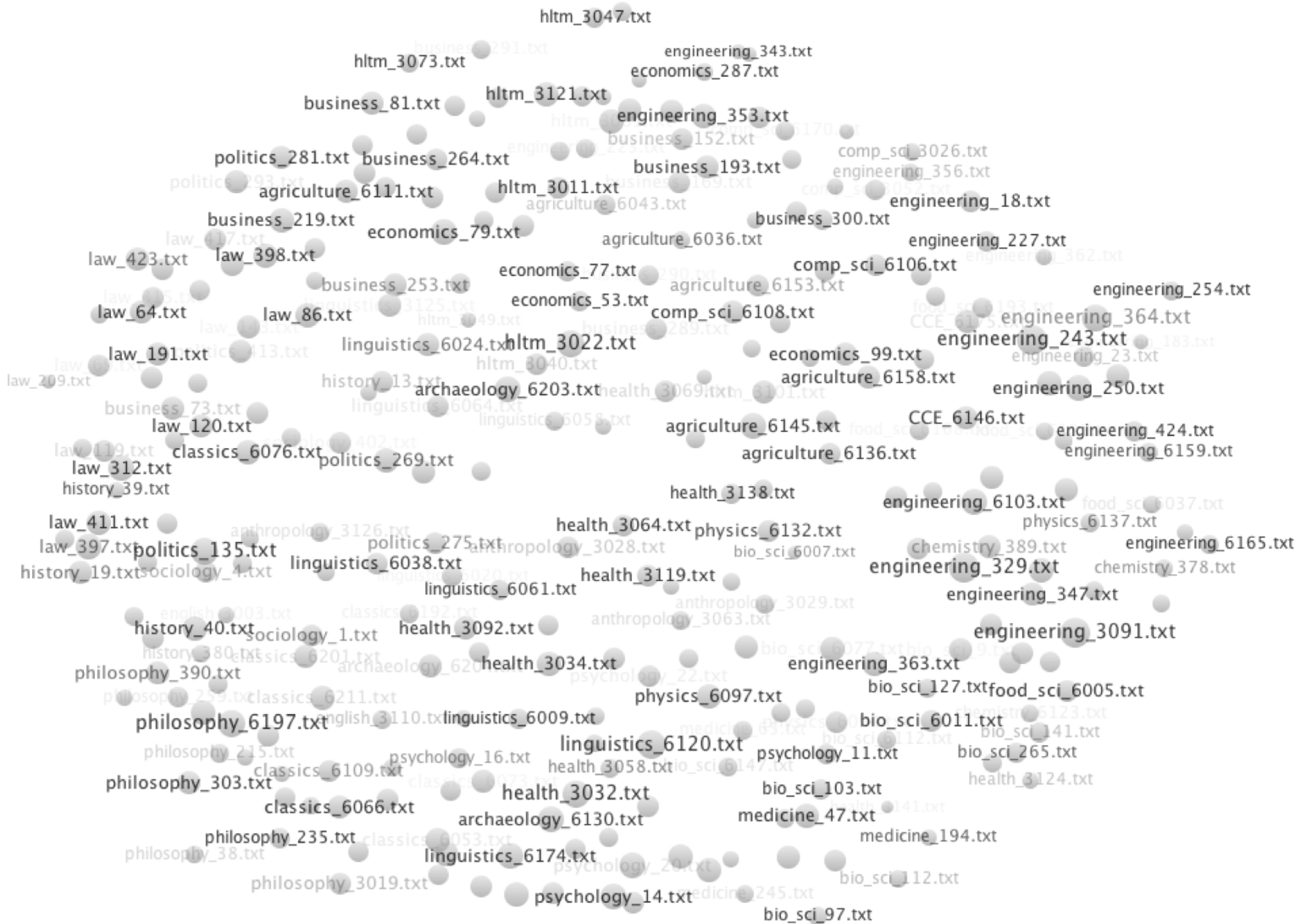
Exploring variation

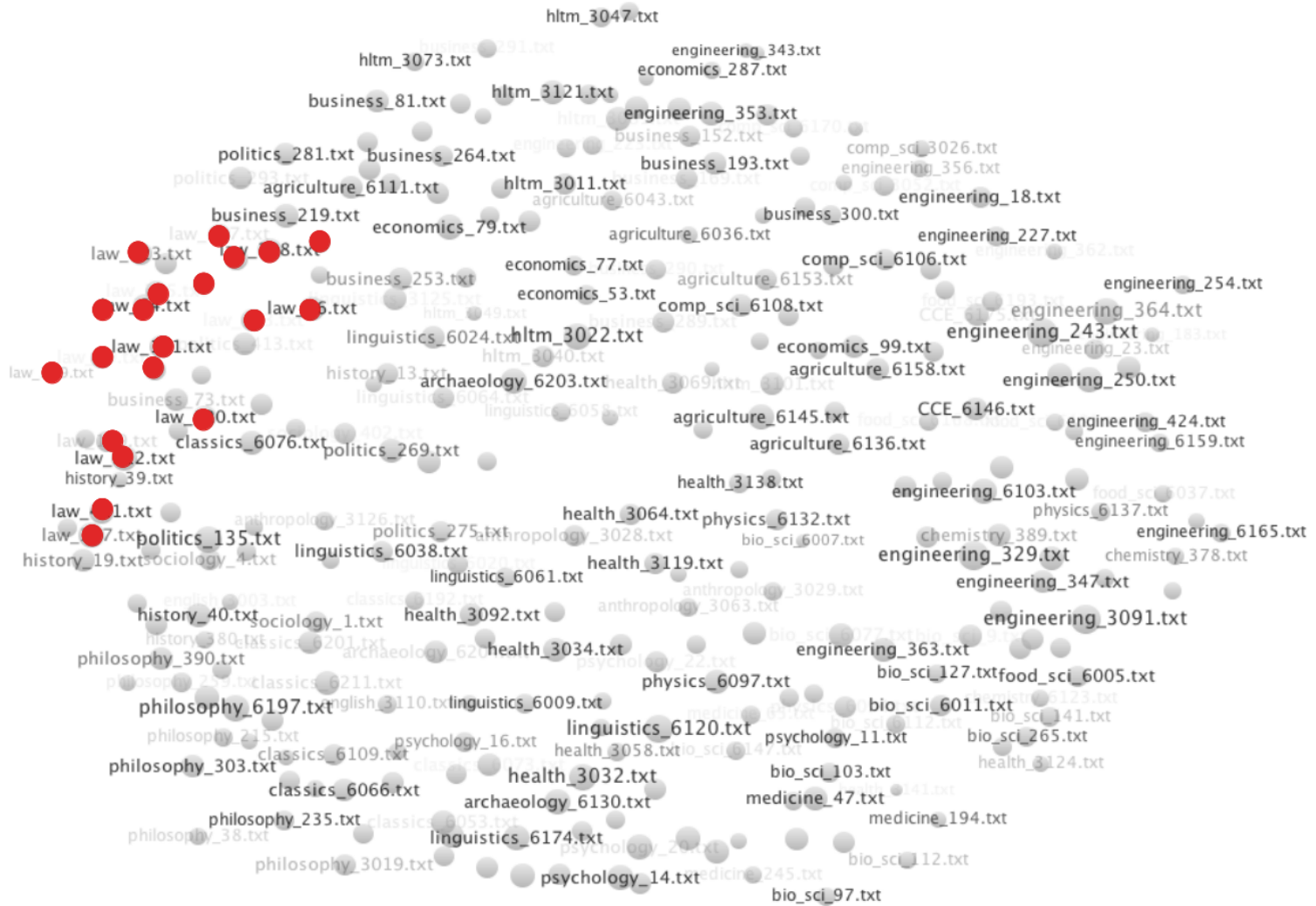
- How do disciplines relate to each other?
 - Borders?
 - Overlaps?
- Are there emergent alternate groupings?
- VOS viewer (van Ek & Waltman 2010)





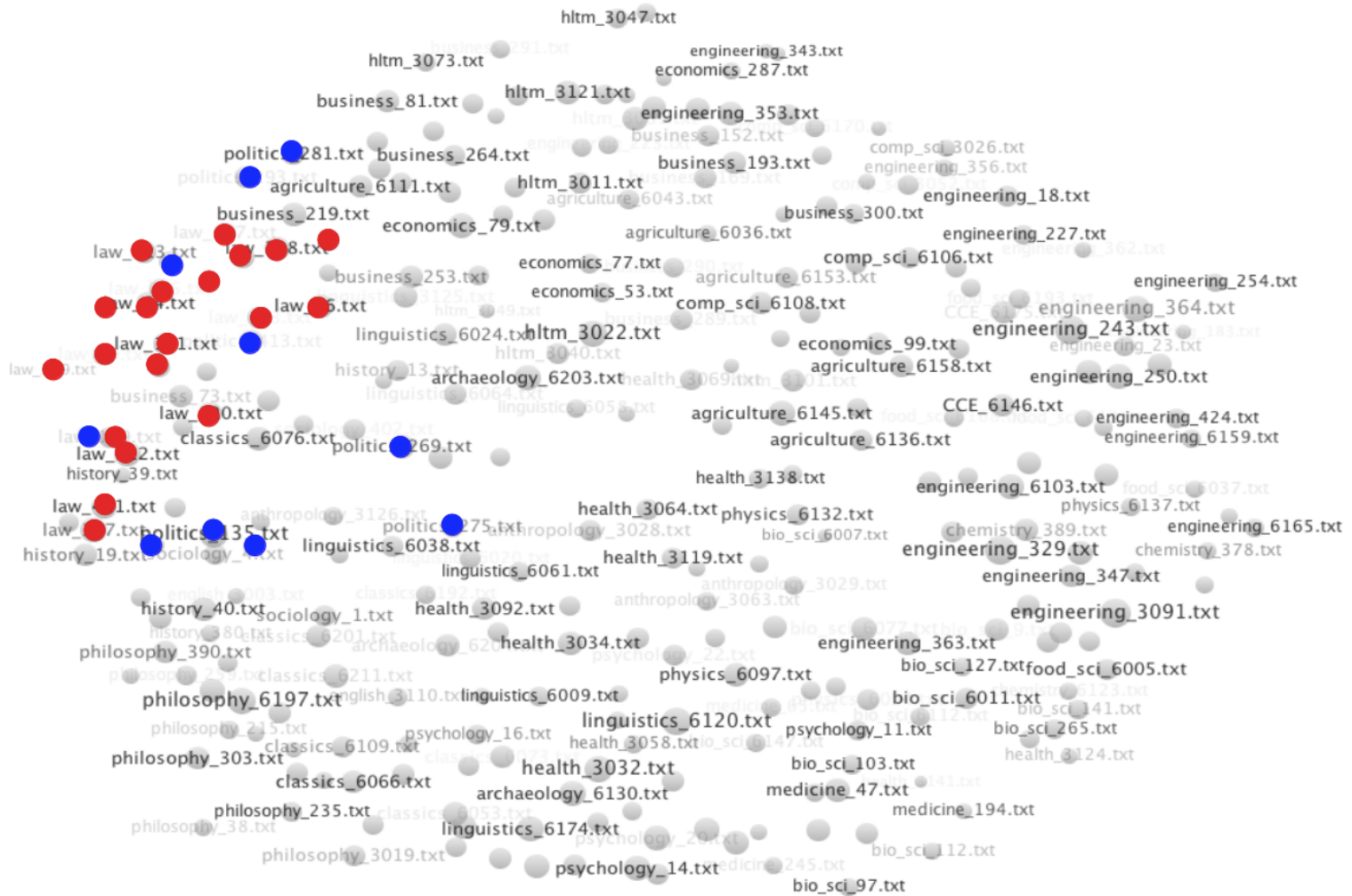






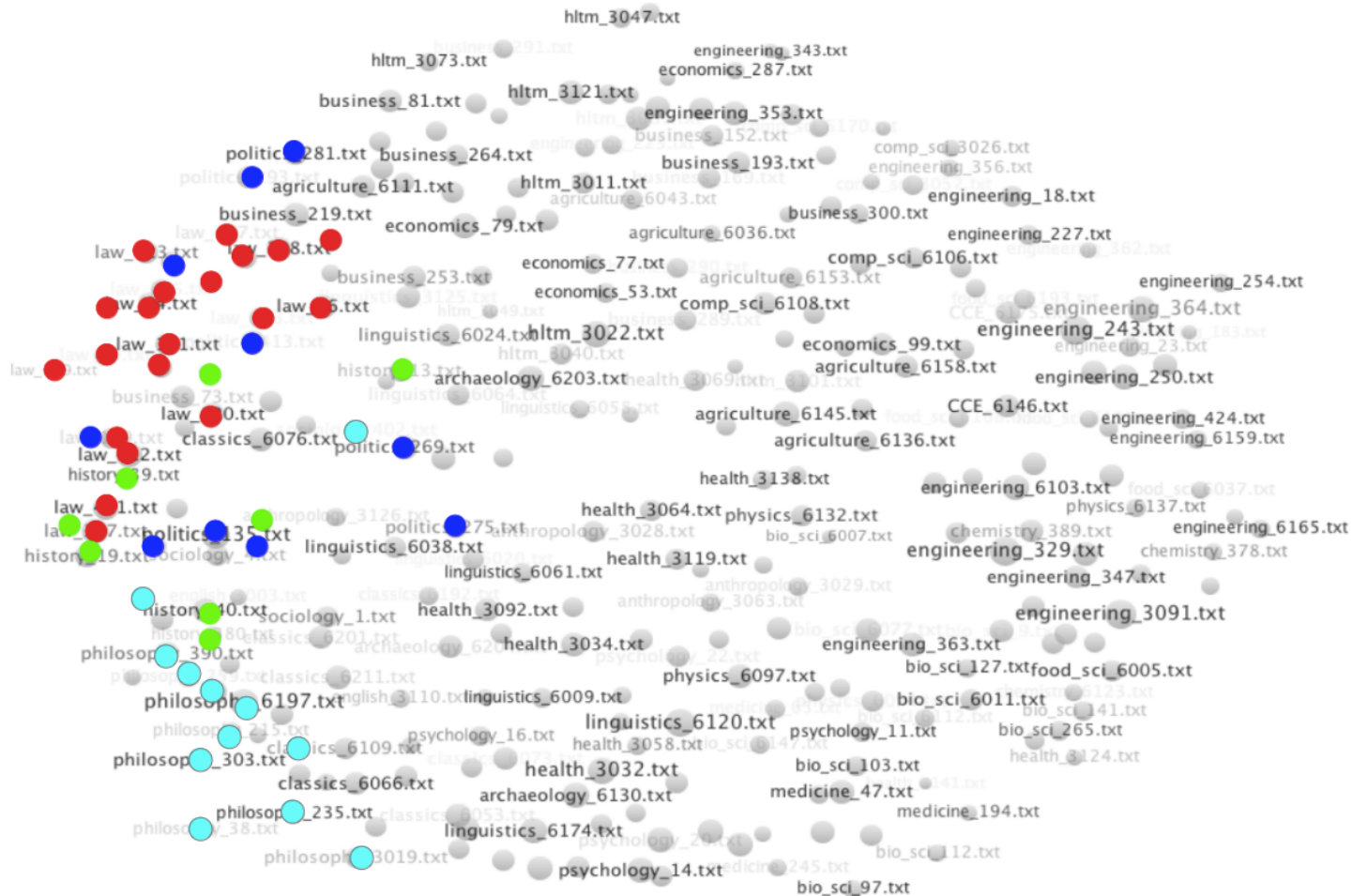
● Law





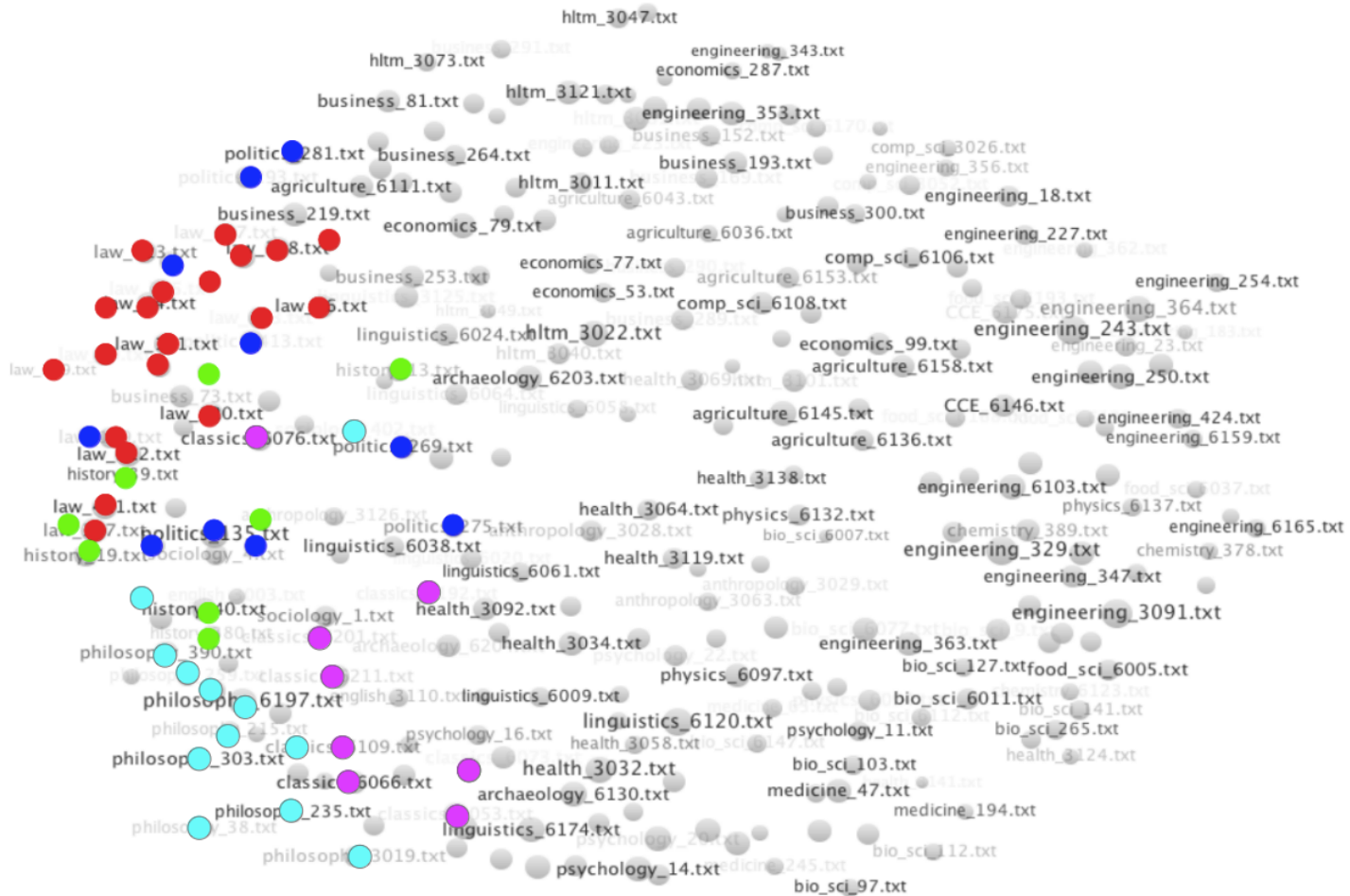
● Law ● Politics





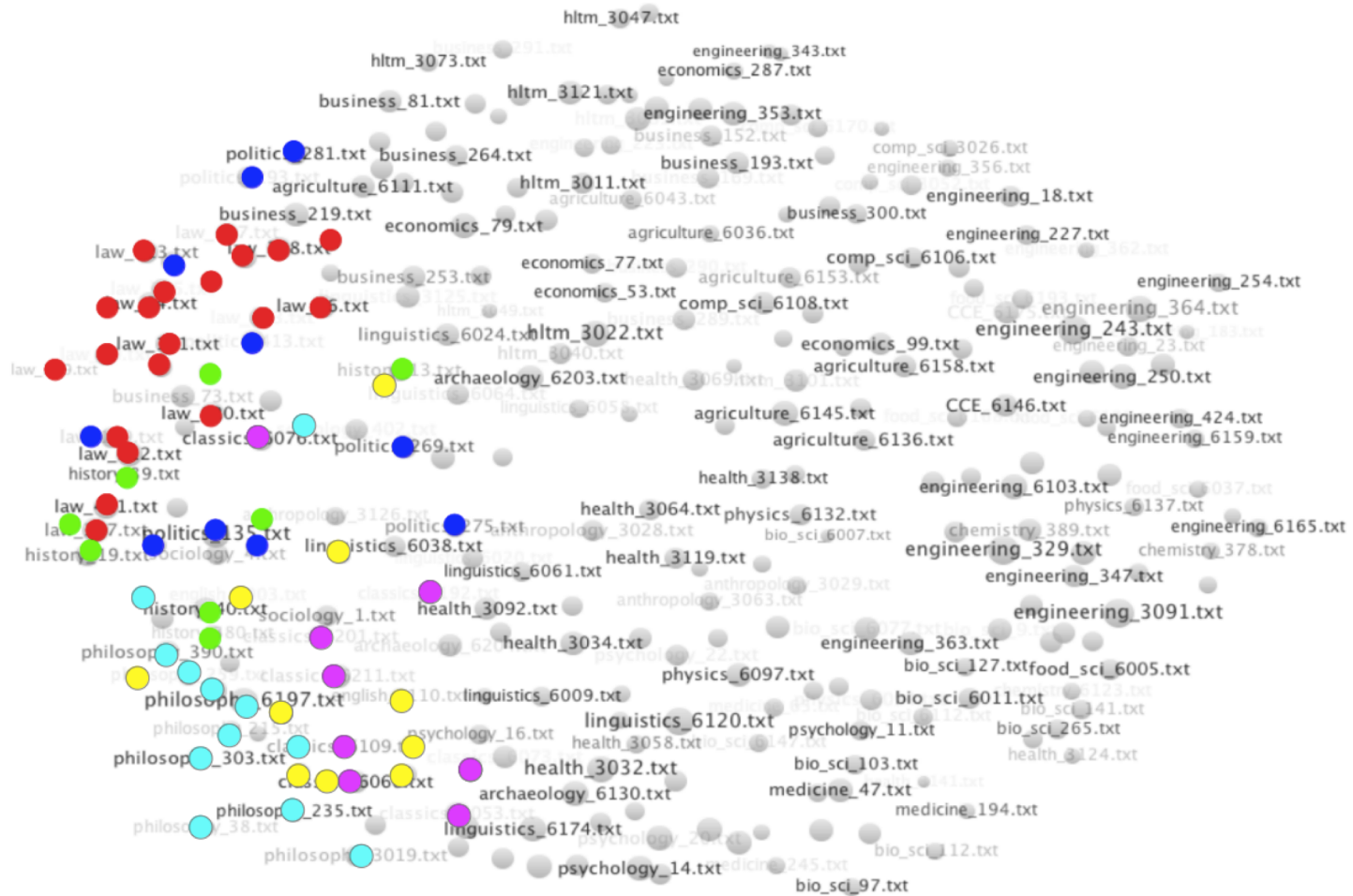
● Law ● Politics ● History ● Philosophy





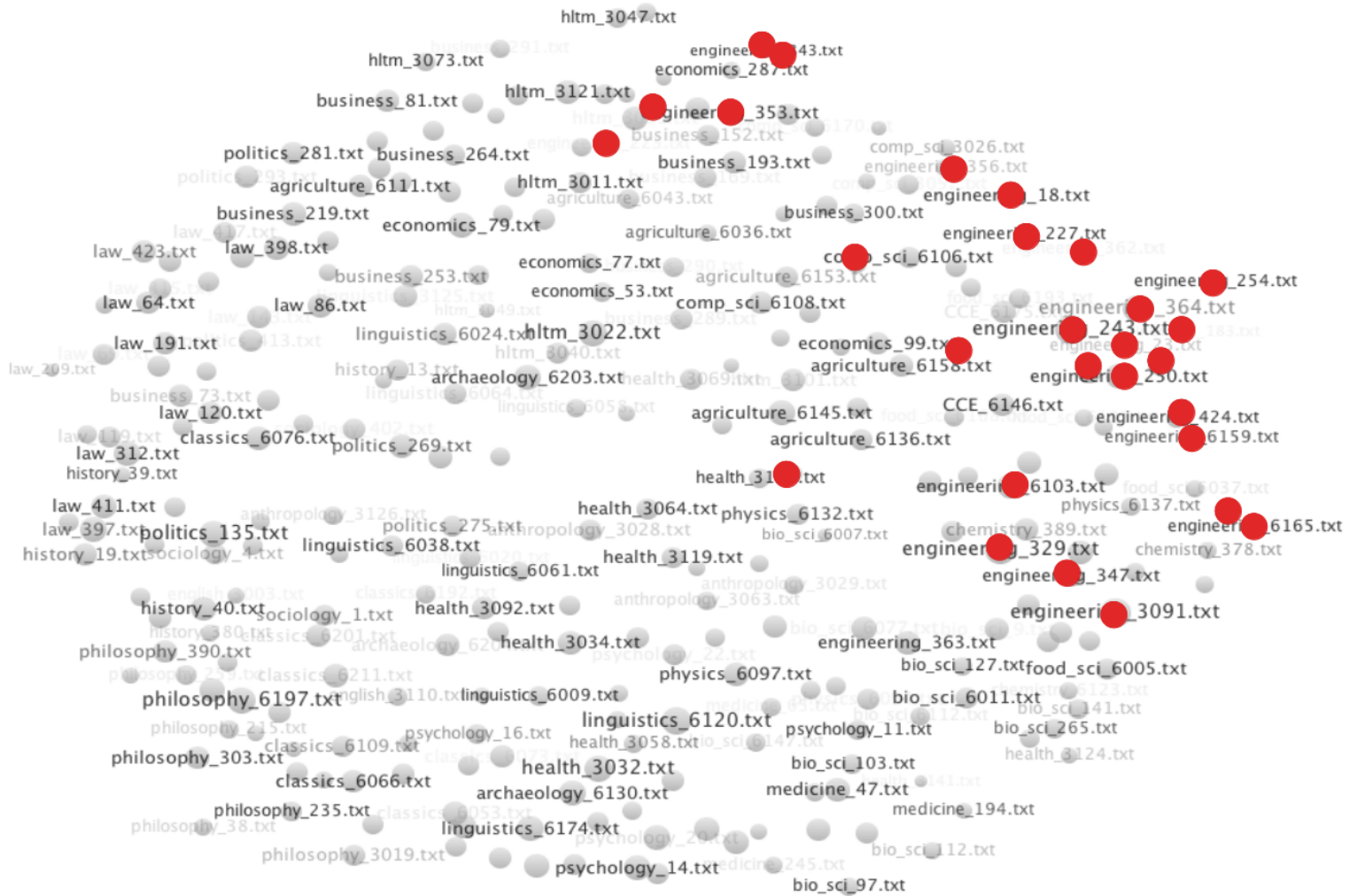
● Law ● Politics ● History ● Philosophy ● Classics





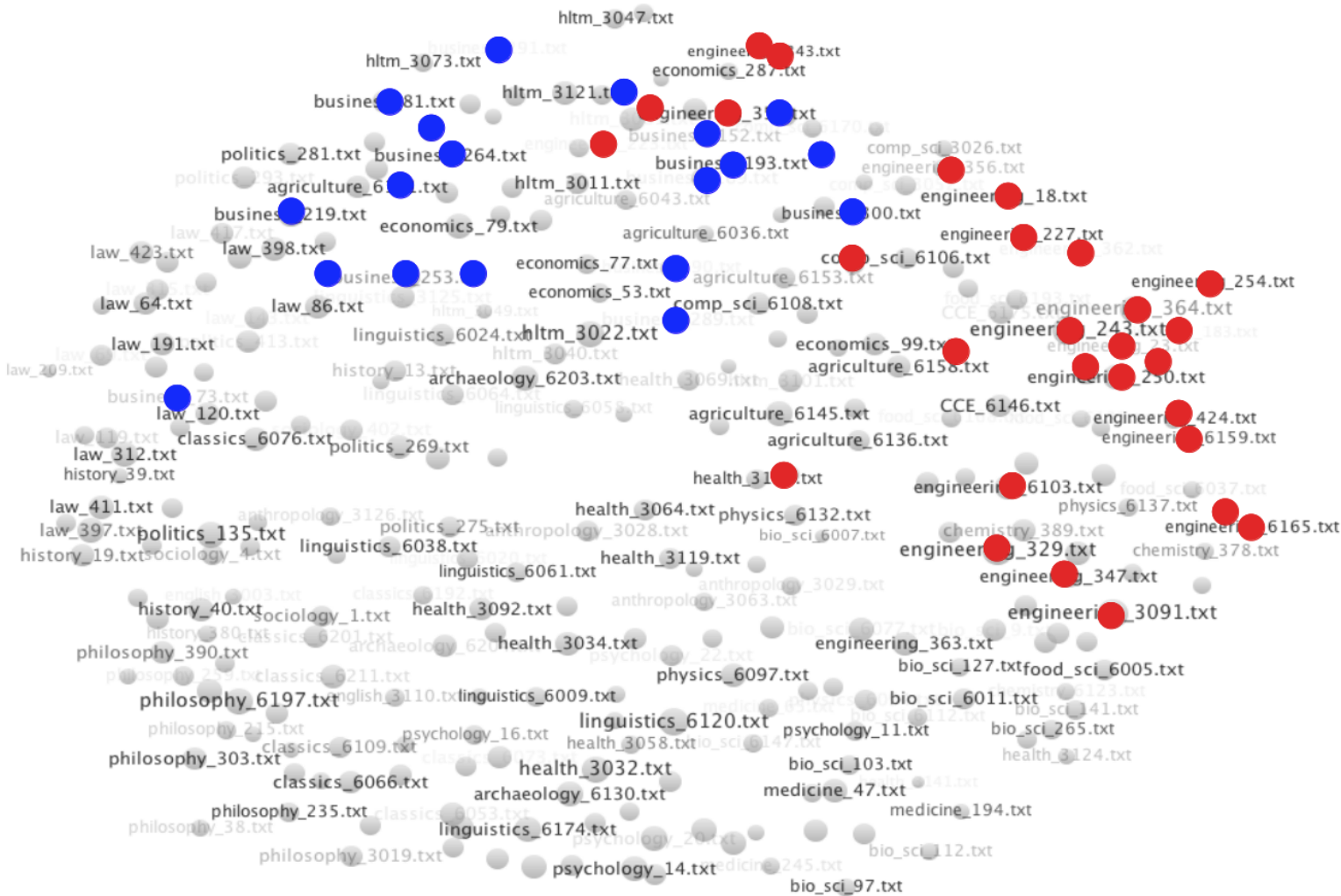
● Law ● Politics ● History ● Philosophy ● Classics ● English





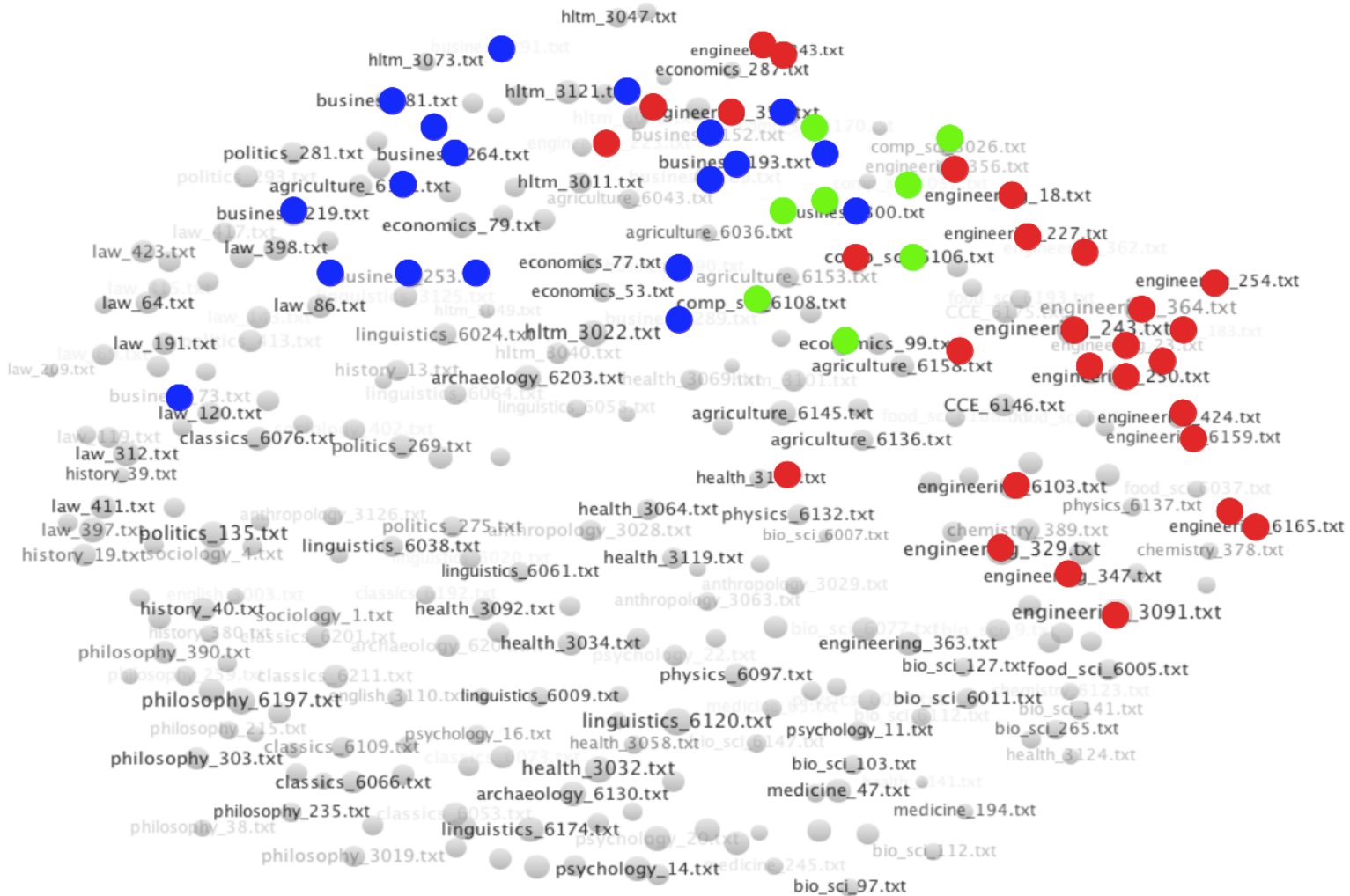
● Engineering





● Engineering ● Business





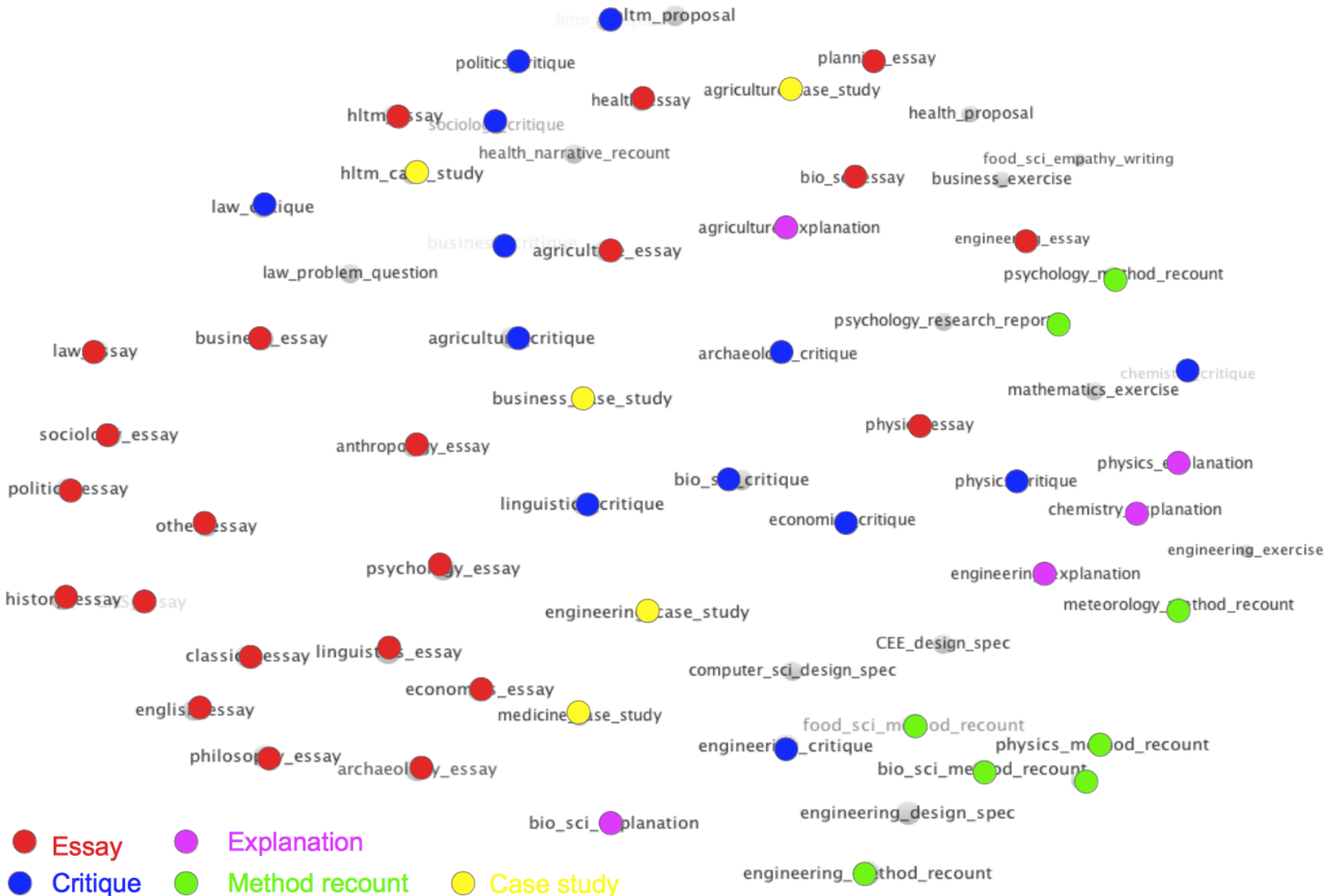
● Engineering ● Business ● Computer Science



Discipline x Genre

	Biological Sciences	Business	Chemistry	Economics	Engineering	English	History	Law	Medicine	Philosophy	Physics	Psychology
Essay	0.15	0.75	0.15	1.27	0.10	1.91	2.19	1.43	0.25	2.06	0.39	1.34
Case study	0.00	3.03	0.32	0.15	1.95	0.00	0.00	0.11	11.93	0.00	0.00	0.00
Research Report	2.12	0.30	0.49	3.24	0.78	0.00	0.00	0.00	0.00	0.00	4.53	5.56
Methodology Recount	2.46	0.11	4.78	0.88	3.05	0.00	0.00	0.00	0.00	0.00	2.21	0.88
Design Specification	0.00	0.41	0.00	0.00	3.72	0.00	0.00	0.00	0.00	0.00	0.89	0.32





Science & Technology vs. Humanities & Social Sciences

- Examine 'distinctive' bundles, i.e.:
 - Wide-spread within the cluster (used by $\geq 20\%$ of writers)
 - Distinctive (\geq twice as widely-used as in other cluster)



Bundle classification

- Research-oriented:
 - Location; Procedure; Quantification; Description; Intangible framing
- Text-oriented:
 - Transition; Resultative; Structuring; Framing
- Stance:
 - Centrality; Epistemic; Modality

Cf. Hyland 2008; Biber et al 2004



Research: Objects of Focus

- Physical world vs. intangibles
 - *the shape of the; on the surface of; the presence of a*
 - *of the concept of; the fact that it; the role of the*
- Quantification:
 - *the length of the; is the same as; the difference between the*
 - *the extent to which*



Research: Procedures & Processes

- Instruments and agents
 - *if stem cells were **to be used in the treatment of diabetes***
(Biological Sciences)
 - ***The role of the West** became increasingly important...*
(History)



Research: Procedures & Processes

- Why things are done
 - *tests should also be conducted **to ensure that the bacteria...*** (Biological Sciences)
 - *...employing these categories **as a means of injecting a bit of order (albeit artificial) into the vast amount of literature*** (History)



Research: Types and Individuals

- *The performance of a collector system is reduced by **the use of a heat exchanger** (Engineering)*
- *...**the use of the** women's own perspective and descriptions, written in their own dialect, allows the reader to...(Politics)*



Stance: Epistemic

- Findings and received opinion vs. points of view and arguments:
 - *it is thought that; it was found that*
 - *be seen as a; appears to be a; it could be argued*



Stance: Evaluation

- *Florence is certain to remain **at the heart of** any serious study of...(History)*
- *It has been **at the heart of** the Single European Market project...(Law)*



Text: Science & Technology

- Cause and effect:
 - *This is due to the fact that as the ph increases there are less hydrogen ions in the buffer* (Biological Sciences)
- Data:
 - *As shown in figure 9 the Vegard line...*(Physics)



Text: Humanities & Social Sciences

- Contrasts:
 - *The term 'phenomenology' refers back to Kantian terminology, where the phenomenon is taken as reality-appearing-to-consciousness, **as opposed to the noumenon** (Philosophy)*
- Framing:
 - *...should be looked at in turn **in relation to the actions and intentions of the two defendants** (Law)*



Text: *this means that the*

- *Let's now assume that the inflation in the country is indeed negative, i.e. the country is facing deflation. **This means that the** real interest rate in the country is positive...(Economics)*
- *Behind this there is either a small backlight or alternative mirror to provide light for the screen to make use of to convey images. **This means that the** thickness of LCDs is small enough that...(Physics)*



Text: *does not mean that*

- *Chinese has no grammatical tense system but this **does not mean that** the speakers have no concept of time...(Linguistics)*
- *Having said that there are modern ideas in Hobbes, this **does not mean that** his end result is a modern state (Sociology)*



Conclusions

- Disciplines are relatively coherent units
- Particular disciplines cross boundaries in interesting ways
- 'Hard' vs. 'soft' is a central axis of variation



Conclusions: Hard disciplines

- Focus on the physical world
- Emphasize the roles of passive, interchangeable instruments which are controlled by researchers
- Quantification
- Received knowledge
- Cause and effect



Conclusions: Soft disciplines

- Focus on abstract concepts
- Emphasize the roles of autonomous agents in processes that are difficult to control
- Multiple contingent viewpoints; evaluation; centrality
- Setting ideas in relation/interpretive contexts



References

- Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpus*, 4(1), 71-83.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Buckingham: SRHE and Open University Press.
- Biber, D. (2006). *University Language*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3), 371-405.
- Brew, A. (2008). Disciplinary and interdisciplinary affiliations of experienced researchers. *Higher Education*, 56: 423-438.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.



References

- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for Academic Purposes. *English for Specific Purposes*, 28(3), 157-169.
- Durrant, P. (2014). Discipline- and level-specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328-356.
- Durrant, P. (in press). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*.
10.1093/applin/amv011
- Gardner, D. & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Simpson-Vlach, R. & Ellis, N.C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4): 487-512.
- Van Eck, N.J. & Waltman. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523-538



THANK YOU

