

The Fortunes of *Philology*

philology, Am. Her. 3d edition

1. Literary study or classical scholarship.
2. See historical linguistics.

An intermediate definition: Cf A. Darmester's "la vie [l'avis] des mots étudiée dans leurs significations" (1886)

The Persistence of Philology

But cf the persistence of philological methodology in intellectual history, political theory, cultural studies, semiotics, literary study, etc.

E.g., work by A. J. Greimas, Raymond Williams, J. G. A. Pocock, Quentin Skinner... and Foucault.

Cf also the “philological” turn in recent philosophy of language -- the emphasis on language as a social fact; the concern with mechanisms of meaning transmission (e.g., Kripke on proper names) and the “linguistic division of labor” (Putnam).

New Tools for Philological Research:

Varieties of Online Corpora

- Regularly updated newspapers & press databases (Nexis, Dialog, Proquest, newspaperarchive.com, America's Newspapers, etc.)
- Standard Web search engines (e.g., Google, AltaVista) and group postings (also Amazon text search)
- Online book corpora from Google, Internet Text Consortium (Internet Archive, Google,
- Scholarly archives: e.g., JSTOR, etc.
- Historical archives: Building of America, Newgate Chronicles, Early American Newspapers, Women's Letters and Diaries, African-American Newspapers, etc. etc.
- Literary archives: ECCO (Eighteenth-Century Corpus Online), Electronic Text Center at UVa, Project Gutenberg, Chadwyck-Healy collections, ARTFL (*Trésor de la langue française*), etc.
- Specific author & publication archives

What Online Corpora Enable

Antedating & Supplementing Dictionaries

E.g., *protest march* (1961/1912)/*protest literature*; *perfidious*; *parlo(u)r pink* (1960/1924), etc.

bleeding heart (OED, 1958): "Praise for what Mr. Pegler calls the 'bleeding heart type' of journalists was heaped upon these broad-minded leaders and a great campaign to cleanse organized labor of the Communistic taint was proclaimed." WSJ, 6/23/1941

Documenting unrecorded collocations; e.g., *traditional* of grammar, weddings, houses, families, etc.

'I place my reliance on your improved sense of duty, and more correct feeling, and better reflection, Madam,' said Mr. Dombey.

What Online Corpora Enable

Antedating & Supplementing Dictionaries, cont.

“Antedating the OED from the JSTOR Electronic Journal Archive” (Fred Shapiro, Yale Law School, funded by Mellon Foundation)

double standard (OED 1951) 1912 *Amer. Polit. Sci. Rev.* VI. 652 Mr. Hecker urges that the double standard of morality for the sexes must gradually be abolished.

Also *counterculture* (OED 1970/1951); *functionalism* (OED 1914/ 1894); *privatize* (OED 1969, 1940); *ambivalence* (OED 1924/ 1914); *role model* (OED 1957/1944), *affirmative action* (OED 1935/ 1906), *recession* (OED 1929/1894)

(Need for care interpreting these)

What Online Corpora Enable

Documenting unrecorded collocations

e.g., *traditional* of grammar, weddings, houses, families, etc.

Correct in Victorian English:

‘I place my reliance on your improved sense of duty, and more correct feeling, and better reflection, Madam,’ said Mr. Dombey. (Ch 40)

‘The first Mrs. Dombey had great good sense,’ said Mr. Dombey, in a gentlemanly toleration of the dead, ‘and very correct feeling.’ (Ch 42)

He certainly rose greatly in my esteem within the last two hours, simply by the propriety of his manners, and the degree of correct feeling with which he had, without any ostentation, coupled their exhibition. W. G. Simms, *Richard Hurdis*, ↓ (1838)

What Online Corpora Enable

Concordances on the fly

Cf Jane Austen's *man of information* (Rachel Brownstein, "Jane Austen and the Dawn of the Information Age")

"Is he a man of information?" (Emma to Jane Fairfax, of Frank Churchill)

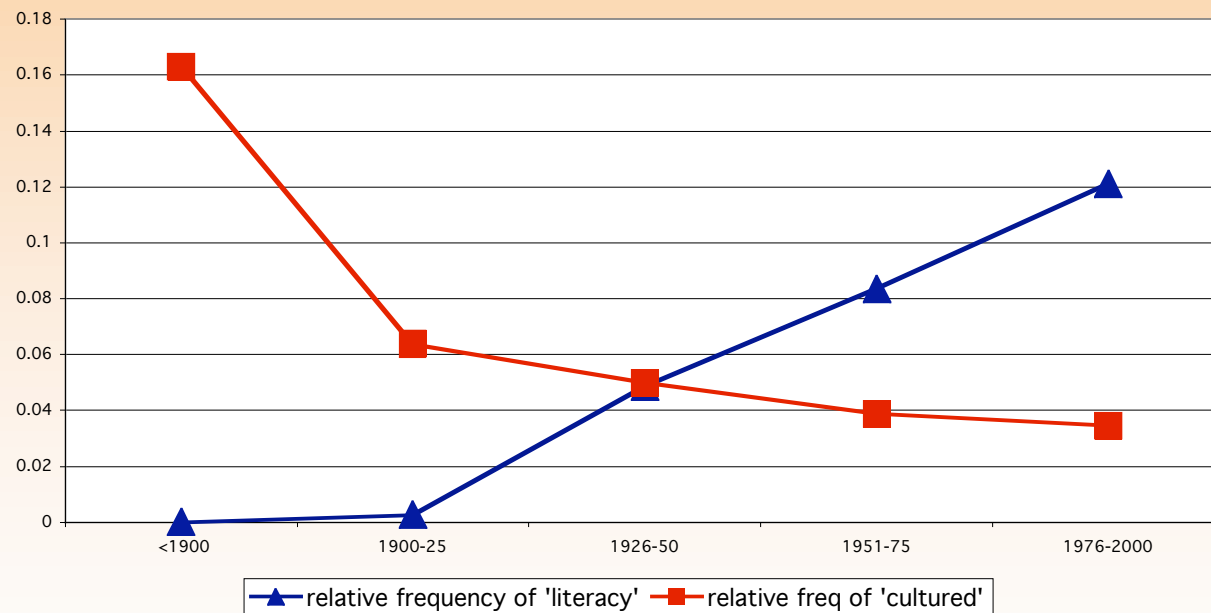
"Mr. Martin, I suppose, is not a man of information beyond the line of his own business. He does not read?" (Emma to Harriet Smith)

"[Susan was] without any of the early delight in books which had been so strong in Fanny, with a disposition much less inclined to... information for information's sake..." (JA of Fanny Price's sister Susan, in *Mansfield Park*)

"Where any one body of educated men, of whatever denomination, are condemned indiscriminately, there must be a deficiency of information, or (smiling) of something else." (Edmund Bertram to Mary Crawford in *Mansfield Park*)

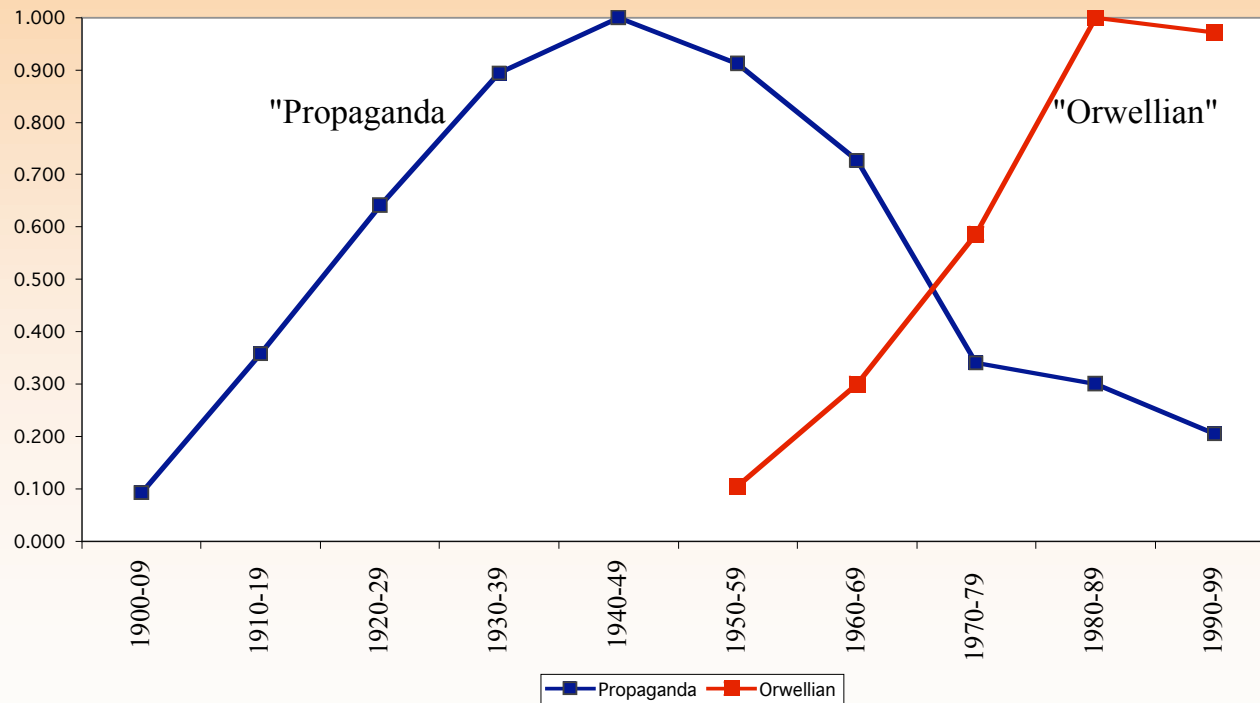
Tracking Change in Frequency

Relative Frequencies of "literate" and "cultured"
in JSTOR Journals in Education and Language and Literature
(frequencies normalized for db size)



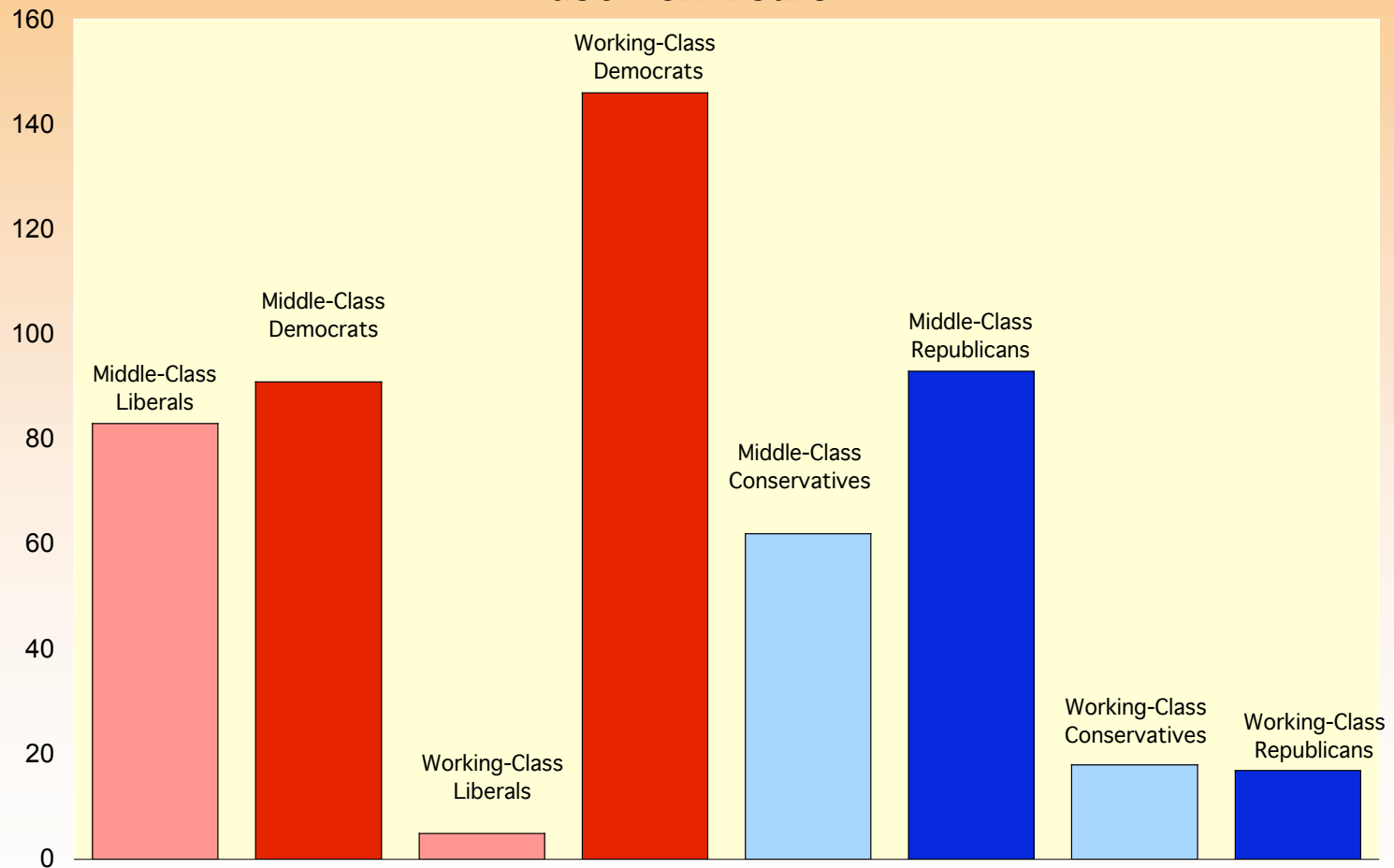
Tracking Historical Change

Relative Annual Frequency of "Propaganda" and "Orwellian" in The New York Times, by Decade (Normalized)



Disparities in Frequency

Descriptions of Voters in Major Newspapers,
Past Ten Years



Teasing Out Semantic Distinctions

in the circumstances: 3,310,000/under the circumstances: 3,980,000
(in/under ratio: 0.83)

in reduced ~: 24,600/under reduced ~ : 174 (ratio: 141.38)

in your ~ : 98,400/under your ~ : 1,940 (ratio: 50.72)

in what ~ : 318,000/under what ~ : 1,680,000 (ratio: 0.19)

in those ~ : 1,070,000/under those ~ : 646,000 (ratio: 1.66)

in those ~ where: 74,100/under those ~ where: 538 (ratio: 137.73)

in many ~ : 298,000/under many ~ : 48,500 (ratio: 6.14)

in a few ~ : 21,300/under a few ~ : 462 (ratio: 46.10)

in no ~ : 289,000/under no ~ : 6,260,000 (ratio: 0.05)

(Thanks to Arnold Zwicky)

Some Practical Issues

Linguists/lexicographers etc. often have different requirements from literary or historical researchers...

- number of hits returned

- formats of documents returned

- availability of kwic (contrast Gutenberg and UVa)

- amenability to automated routines

- availability of stemming (or sometimes NOT)

- full Boolean search, search on metadata, etc.

- agglomeration of collections for metasearch...

“Occurrent” corpora are rarely designed with linguistic purposes in mind... nor do they always accommodate this interest in cost-structures... (though sometimes, accidentally, they do).

Some Practical Issues, 2

Data providers are not always interested in working to standardize features, enable metasearch, etc.

Researchers are often hampered by unfamiliarity with resources, tools, techniques, or interfaces.

Resources are often costly or inaccessible, particularly if they are regarded as essential scholarly infrastructure.

Some Theoretical Issues

What is the relation between words and social concepts? I.e., what can we conclude from the appearance of a new word or new sense?

Cf Williams vs. Skinner on *patriot*; advent of *protest* as a mass term, earliest use of *religious* to mean "conscientious," etc.