

humanities computing

A short history of the “first phase” in
10½ slides.

memory

A simultaneous possession by the reader of all the words and images of *Middlemarch*, *À la recherche du temps perdu*, or *Ulysses* may be posited as an ideal, but such an ideal manifestly cannot be realized. It is impossible to hold so many details in the mind at once.

J. Hillis Miller, "Three Problems of Fictional Form: First-person Narration in David Copperfield and Huckleberry Finn." In R. H. Pearce (ed.), *Experience in the Novel: Selected Papers from the English Institute*. New York: Columbia University Press 1968, 23.

memory & the critic

It is a truth not generally acknowledged that, in most discussions of works of English fiction, we proceed as if a third, two-fifths, a half of our material were not really there.

John Burrows, A Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method. Oxford 1987, 1.

computers & humans

Seen as a mere tool without any inherent analytical power of its own, the computer in literary studies enhances the critic's powers of memory electronically, thereby providing a complete database of findings that meet all predefined patterns or search criteria.

Thomas Rommel, "Literary Studies." In S. Schreibman, R. Siemens, J. Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell 2004, 89.

computers & humans

In order to be successful, literary computing needs to use techniques and procedures commonly associated with the natural sciences and fuse them with humanities research, thereby bringing into contact the Two Cultures [...].

Thomas Rommel, "Literary Studies." In S. Schreibman, R. Siemens, J. Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell 2004, 89.

concordance

An alphabetical arrangement of the principal words contained in a book, with citations of the passages in which they occur. These were first made for the Bible; hence Johnson's explanation 'A book which shows in how many texts of scripture any word occurs'.

Orig. in *pl.* (*medieval Latin concordantiæ*), each group of parallel passages being properly a *concordantia*. This is sometimes denominated a *verbal concordance* as distinguished from a *real concordance* which is an index of subjects or topics.

OED

concordance

- index
- lists all words alphabetically, with occurrences
- around 1230 first concordance
- *Vulgate* (500 Dominican monks)
- Hugo de Saint-Cher, printed 1470 (Latin)

concordance

- Alexander Cruden (1699-1770), bookseller
 - King James Bible (1604-11)
 - singlehandedly in one year
 - published 1737
 - never out of print
- James Strong (1822-1894), Prof. of Theology
 - King James Bible
 - >100 contributors
 - published 1890
 - Strong's number: original language

KWIC: key word(s) in context

- function words vs. content words
- lemmatization
- disambiguation
- variant spelling, typos
- synonyms

kwic

- offline concordance Adam Smith, *The Wealth of Nations*, [adobe reader](#)
- online concordance William Blake, *Collected Works*, [web interface](#)

Roberto Busa

- Father Roberto Busa (Jesuit priest, 1913-)
- 1949 index verborum of Thomas Aquinas (*Index Thomisticus*)
- 11 million words
- printed concordance 1974
- Roberto Busa Award [ADHO](#) (Alliance of Digital Humanities Organizations)

markup

- angle brackets <tag> arbitrary choice
- COCOA concordance program (1967)
- idiosyncratic, non-standardized
- first SGML, now XML
- TEI: Text Encoding Initiative

style

- individual choice of syntax and vocabulary (cf poetic function)
- not always under author's control
- “style markers”
- distribution
- frequency (hapax legomena)
- function words

style

- repetition (redundancy)
- type-token ratio
- “Rose is a rose is a rose is a rose”
- type: abstract (3 types of words)
 - a
 - rose
 - is
- token: example (10 words in total)

stylometrics

- authorship studies:
 - who wrote Shakespeare? Who wrote Anonymous
 - # of words of two/three letters in Shakespeare and contemporaries
 - punctuation
 - word length
 - sentence length
 - word distribution
 - function words
 - content words
 - type-token ratio
- forensic studies
- plagiarism

authorship studies

- Alvar Ellegård (1962): Junius Letters
- Andrew Morton (1963): epistles by St. Paul (frequency of common words)
- Fred Mosteller & David Wallace (1964): identify author of *Federalist Papers* (US constitution)
- Centre for Linguistic and Literary Computing in Cambridge, TUSTEP at Tübingen University
- Computers and the Humanities (1966)

common applications

- literary-linguistic studies (stylistics)
 - style: repetition, omission, syntax, word usage
 - prosody
- thematic studies
 - intertextual references (cf. plagiarism)
 - presence/absence of concepts
- translation studies
 - parallel versions
 - vocabulary
- editing and editions
 - notes, apparatus, sources
 - dynamic (incremental) editions
 - meta editions (inclusion/exclusion documented in anthologies)

David Lodge, *Small World* (1984)

"*What's the use?* Let's show him, Josh."

And he passed the canister to the other guy, who takes out a spool of tape and fits it on to one of the machines.

"Come over here", says Dempsey, and sits me down in front of a kind of typewriter with a TV screen attached.

"With that tape", he said, "we can request the computer to supply us with any information we like about your idelect."

"Come again?" I said.

"Your own special, distinctive, unique way of using the English language. What's your favourite word?"

"My favourite word. I don't have one."

"Oh yes you do!" he said. "The word you use most frequently."

David Lodge, *Small World* (1984)

That's probably *the* or *a* or *and*,' I said.

He shook his head impatiently.

'We instruct the computer to ignore what we call grammatical words—articles, prepositions, pronouns, modal verbs, which have a high frequency rating in all discourse. Then we get to the real nitty-gritty, what we call the lexical words, the words that carry a distinctive semantic content. Words like *love* or *dark* or *heart* or *God*. Let's see.'

So he taps away on the keyboard and instantly my favourite word appears on the screen. What do you think it was?