Authors & the Classification of Discourse

Geoff Nunberg
Quality of Information: 10/4/06
The Notion of "Discourse"

Discourse -- a faux-theoretical term?

In linguistics & sociology of language: structures of language above the sentence.

Cf issues involving pronoun reference, discourse connectives, turn-taking, relevance, etc.
Discourse in literary theory:
Instead of gradually reducing the rather fluctuating meaning of the word ‘discourse’ I believe I have in fact added to its meanings: treating it as sometimes the general domain of all statements, sometimes as an individualisable group of statements, and sometimes as a regulated practice that accounts for a number of statements. Michel Foucault, *The Archeology of Knowledge*, 1972

E.g.: "medical discourse," etc.

Discourse as "episteme"/institutionalized rules of thought
Discourse sets "the limits of acceptable speech" (Judith Butler)
The Notion of "Discourse"

Leaving well enough alone

Discourse = the ambient yammer & its associated structures
In common usage, discourses can be individuated by:

Field/subject ("medical discourse," "political discourse")
Community/speakers ("adolescent discourse," "classroom discourse")
Purpose ("instructional discourse," "bull sessions")
Mode of communication ("computer-mediated discourse," "spoken discourse")

Etc.

Discourse as distinct from mode & genre (more later)

  e.g. discourse ≠ novels, comedy, etc.
The Ontologies of Discourses

Relative to a particular discourse, we distinguish an interconnected set of relevant actors, units, classes, relations, rules of composition, principles of interpretation, legal & ethical constraints, etc.

E.g. in "classical" literary discourse:

- works, œuvres, authors, schools, genres, notions of plagiarism, etc.

In scientific discourse:

- publications, "authors," genres, rules of citation, etc.
The Notion of the Author

Authors & writers:

USAGE NOTE: The verb *author*, which had been out of use for a long period, has been rejuvenated in recent years with the sense "to assume responsibility for the content of a published text." As such it is not quite synonymous with the verb *write*; one can write, but not author, a love letter or an unpublished manuscript, and the writer who ghostwrites a book for a celebrity cannot be said to have "authored" the creation. The sentence *He has authored a dozen books on the subject* was unacceptable to 74 percent of the Usage Panel, probably because it implies that having a book published is worthy of special lexical distinction, a notion that sits poorly with conventional literary sensibilities and seems to smack of press agentry. The sentence *The Senator authored a bill limiting uses of desert lands in California* was similarly rejected by 64 percent of the Panel, though here the usage is common journalistic practice and is perhaps justified by the observation that we do not expect that legislators will actually write the bills to which they attach their names. • The use of *author* as a verb in computer-related contexts is well established and unexceptionable.
Modern qualms:

Author as a verb:

*He has authored a dozen books on the subject.*
*The Senator authored a bill limiting uses of desert lands in California.*
*She co-authored an important paper on Albanian orthography.*

Authors & writers

- Charles Dickens
- John Keats
- Wm. Shakespeare
- John Updike
- Paul Duguid
- Noam Chomsky
- Stephen King
- Sigmund Freud
- Paul Krugman (Maureen Dowd, etc.)
Authors, works, & œuvres are interdefined.

"At this point, however, a problem arises: What is a work? What is this curious unity which we designate as a work? Of what elements is it composed? Is it not what an author has written? Difficulties appear immediately. If an individual were not an author, could we say that what he wrote, said, left behind in his papers, or what has been collected of his remarks, could be called a 'work'? When Sade was not considered an author, what was the status of his papers?"

But the "author function" is not identified with the historical individual who composed the writings.
Which discourses contain an author function?

Author function characterized by:

Legal codification:

Speeches and books were assigned real authors, other than mythical or important religious figures, only when the author became subject to punishment and to the extent that his discourse was considered transgressive.

Historically specific (e.g., modern conception of "literary" discourse

A constructed rather than natural role

This "author-function" … is not formed spontaneously through the simple attribution of a discourse to an individual.

The author function does not refer simply to a mere individual.
Varieties of Transgression:
The question of plagiarism

Alex Haley, Roots:

Stephen Ambrose, *The Wild Blue*: "Up, up, up he went, until he got above the clouds. . . B-24's, glittering like mica, were popping up out of the clouds. . ." (Thomas Childers)

Doris Kearns Goodwin, *The Fitzgeralds and the Kennedys* "Hardly a day passed without a newspaper photograph of little Teddy taking a snapshot with his camera held upside down, or the five Kennedy children lined up on a train or bus." (Lynne McTaggert)-- sloppy note taking?

"it's plagiarism whether you intended to do so or not."-- American Historical Association spokesperson

"You mean I didn't footnote my references?"
Who Can Plagiarize What?
Who Can Plagiarize What?

Can you plagiarize (and when?) …

a wikipedia article?

a dictionary definition?

a TV listing?
Fixing the text: capturing the "authorial intention"; the notion of a "corrupt" text.
"Author" as a "Classificatory Function"

.. an author's name is not simply an element in a discourse...; it performs a certain role with regard to narrative discourse, assuring a classificatory function. Such a name permits one to group together a certain number of texts, define them, differentiate them, and contrast them to others.

"The author-work relation is embedded in library catalogues, the indexes of standard literary histories.... It is pervasive in our education system ... institutionalized in our system of marketing cultural products ... the name of the author .... becomes a kind of brand name." Mark Rose, *Authors & Owners*

"The name as an individual trademark ..." Foucault
Two levels of classification

(Literary) texts into œuvres
Writings into "authorial" etc.

In a civilization like our own there are a certain number of discourses that are endowed with the "author-function," while others are deprived of it. A private letter may well have a signer -- it does not have an author. A contract may have a guarantor; it does not have an author. An anonymous text posted on a wall probably has a writer; it does not have an author. The author-function is therefore characteristic of the mode of existence, circulation, and functioning of certain discourses within a society.
Classical vs scientific

In the 17th or 18th century, "Scientific discourses began to be received for themselves, in the anonymity of an established or always redemonstrable truth... The author-function faded away, and the inventor's name served only to christen a theorem, proposition, particular effect... or pathological syndrome."

Cf Wm. Whewell (1837) on the way science absorbs its own past:

Previous doctrines may require to be... expressed in new language, to be taken into the body of science by various processes; but ... nothing which was done was useless or unessential, though it ceases to be conspicuous and primary.
The word *literature* is a perpetual source of confusion, because it is used in two senses, and those senses liable to be confounded with each other. In a philosophical use of the word, Literature is the direct and adequate antithesis of Books of Knowledge. But, in a popular use, it is a mere term of convenience for expressing inclusively the total books of a language. In this latter sense, a dictionary, a grammar, a spelling-book, an almanac, a pharmacopœia, a Parliamentary report, a system of farriery, a treatise on billiards, the Court Calendar, &c., belong to the literature. But, in the philosophical sense, not only would it be ludicrous to reckon these as part of the literature, but even books of much higher pretensions must be excluded—as, for example, books of voyages and travels, and generally all books in which the matter to be communicated is paramount to the manner of form of its communication. Thomas De Quincy, "Letter to a Young Man whose Education has been Neglected," (1823)
Other classificatory principles:
The Notion of Genre

Classical sense: a type of text defined by content, language, and form.
E.g., Elizabethan sonnet

Fictional genres

I met Bobby Callahan on Monday of that week. By Thursday, he was dead. He was convinced someone was trying to kill him and it turned out to be true, but none of us figured it out in time to save him. I've never worked for a dead man before and I hope I won't have to do it again.

What kind of text is this?
What do we know about the narrator (married/single, etc.), plot, nature of crimes, resolution, clues, etc. (Who is "us"?)
The notion of "generic cues"
Genre as rules of interpretation
The Functions of Genre

The purposes of Genre

The uses of uniformity

  Cognitive efficiency: reduce possible number of readings

  Establish principles of interpretation, validation, authorial responsibility, etc.
Classification of Genres by Agentive Roles

Some types of texts

"authorial texts"

"Informational" texts --
  "signed": scientific & scholarly communications
  "unsigned": reference works (dictionary, train schedule, high-school yearbook);
  public notices, etc.

Mixed: newspapers

Historical mutability of these categories
Connection to mode of validating, etc.
Connection to "the mode of existence, circulation, and functioning…"
Genres of Texts and Documents

Which is the hardcover?
Some Distinguishing Features of Document Genres

Mode of production
   Individual/institutional

Mode of circulation
   published/private
   Occasional vs. Periodical ("daily"/"weekly"/"monthly" etc.)

Physical form
   books, newspapers, pamphlets, newsletters, magazines, etc.

Other features?
Problems of Characterizing Digital Genres

How to characterize genres when differences in material form and mode of distribution are no longer present or altered?

Legacy labels: "online journal," etc.

Primary & secondary representations; e.g., online versions of print journals
Problems of Characterizing Digital Genres

New Labels: blogs, weblogs, wikis, "news sites" (Wired News etc.)
"The great relabeling" of blogs as "Online magazines."

What happens to principles of responsibility/validation associated with particular genres; issue of reliability, objectivity, etc.?
What defines a blog as genre?

Properties of form
Properties of content
Properties of language

Can we distinguish document & textual properties?

Who fixes properties & how?
Classifying documents in the print world:
The role of the physical properties of the book

As “Information” — by topic

As “literatures” — by social use, popularity, reputation, etc.

As “reading matter” — by genre, format, etc.
Classifying documents in the digital world

In the world of print...

As “Information”—by topic

As “literatures”—by social use, popularity, reputation, etc.

As “reading matter”—by genre, format, etc.

In the digital world...

Keywords, relevance, summarization, etc.
Genre Identification

Object
Identify the genre of a document, e.g., newspaper article, opinion, market analysis
Use easy-to-compute, low-level features: punctuation, word-length, word-shape, etc.

Northrop (with Brett Kessler and Hinrich Schuetze)
Database of 4400 records on AIDS:
- technical articles, book reviews, company profiles, drug data reports,
- financials, market research, medical research, news stories, opinion, patent
- applications

Genre classification performed on output of relevance search
Classifications 92 percent correct
Search words:

heterosexual spread Africa

Sort by Relevance  Sort by Genre

Search only the following genres:
✓ book review
✓ calendar
✓ company profiles
✓ drug data
✓ financials
✓ market research
✓ medical research
✓ news
✓ opinion
✓ patents
Northrop search results

<table>
<thead>
<tr>
<th>Word</th>
<th>Number of Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>heterosexual</td>
<td>234</td>
</tr>
<tr>
<td>spread</td>
<td>574</td>
</tr>
<tr>
<td>africa</td>
<td>513</td>
</tr>
</tbody>
</table>

42 texts satisfy query. (Relevance weight is in parentheses.)

To see these results sorted by genre, go the form at the end of this page, select the genres you are interested in,

1. (2888) AIDS in the global village: why US physicians should care about HIV outside
2. (504) Epidemiology: Current and future dimensions of the HIV/AIDS pandemic
3. (300) Original Articles: Prevalence of human T-lymphotropic retroviruses
4. (260) Preventing the heterosexual spread of AIDS: are we giving our patients the
5. (180) Probability of female-to-male transmission of HIV-1 in Thailand
6. (180) Heterosexual AIDS: pessimism, pandemics, and plain hard facts. (Editorial)
7. (130) HIV-1 subtypes and male-to-female transmission in Thailand.
8. (120) Slower heterosexual spread of HIV-2 than HIV-1
9. (96) Heterosexual transmission of HIV. (includes reply). (Letter to the Editor)
10. (42) A NEW STRAIN OF AIDS FOUND IN CALIFORNIA
Northrop search results

<table>
<thead>
<tr>
<th>Word</th>
<th>Number of Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>heterosexual</td>
<td>234</td>
</tr>
<tr>
<td>spread</td>
<td>574</td>
</tr>
<tr>
<td>africa</td>
<td>513</td>
</tr>
</tbody>
</table>

42 texts satisfy query. (Relevance weight is in parentheses.)

opinion

1 (2888) AIDS in the global village: why US physicians should care about HIV outside
2 (260) Preventing the heterosexual spread of AIDS: are we giving our patients the
3 (180) Heterosexual AIDS: pessimism, pandemics, and plain hard facts. (Editorial)
4 (96) Heterosexual transmission of HIV. (includes reply). (Letter to the Editor)
5 (63) A NEW STRAIN OF AIDS FOUND IN CALIFORNIA
6 (48) THE UGLINESS AROUND AIDS RESEARCH