



Charles Darwin University

Web-based research:

The example of Ebonics

Black, Paul

Published in:

'eNTIER' : the whole works! Proceedings : Northern Territory Symposium 'eNTIER'

Published: 01/05/2000

Document Version

Early version, also known as pre-print

[Link to publication](#)

Citation for published version (APA):

Black, P. (2000). Web-based research: The example of Ebonics. In 'eNTIER' : *the whole works! Proceedings : Northern Territory Symposium 'eNTIER'* (pp. 32-38). Callaway International Resource Centre for Music and Music Education (CIRME).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Web-based Research: The Example of Ebonics

Paul Black, Northern Territory University

The term 'Ebonics' refers to Black English in America, or what some linguists call African American Vernacular English (AAVE). The term was actually coined by Robert Williams in 1975, but it was almost unknown until 18 December 1996. On that date the School Board of Oakland County in California adopted a policy to recognise Ebonics as the first language of most of their African American students, who they thus saw as benefiting from studying standard English as a second language.

In essence this proposal was educationally sound, but some of the wording was incautious and misleading, such as a statement that Ebonics was part of the students' genetic endowment — this sounded like the school board was saying the students were born to speak Black English, when they simply meant that the variety showed traces of an African linguistic heritage.

In any case the School Board's announcement got quite a mixed reception, and far more notoriety than anyone had probably anticipated. Ebonics was soon mentioned in thousands of Web pages, including news articles, scholarly and popular opinion, internet discussions, and many pages of rather racist Ebonics "jokes". While the Oakland policy was soon endorsed by such scholarly groups as the US Centre for Applied Linguistics (1997) and the Linguistic Society of America (1997), it also received considerable ridicule from members of the public who imagined that it meant that Black street jive was going to be taught in California schools under the new and esoteric sounding name of Ebonics.

My paper isn't actually about Ebonics, but rather about what I learned from researching the Ebonics issue on the Web. While the Web has some limitations as a source of information (many are reluctant to make copyright materials available on the Web, for example), it has special potential as a source of data on popular opinion. Aside from Ebonics I've used it to search for opinion and other information on the demise of bilingual education in both California (Black 1999c) and the Northern Territory (Black 1999a). My greatest efforts were devoted to Ebonics, however, through viewing and cataloguing over 1500 Web pages, beginning in mid 1997 and continuing about a year and a half later over the 1998-99 Christmas break.

Black (1999b, 1999d) has reported in some detail on this study of Ebonics sources on the Web. Here I'll simply address the following points briefly:

1. What I found: Types of data
2. How I found it: Data gathering techniques
3. How much is out there?
4. How long does it stay?
5. How good are the search engines?

1. What I Found: Types Of Data

Of the 1508 Web pages I catalogued (in Black 1999d), some could easily be categorised into the following three types:

- (a) News media pages: there were 332 pages that apparently formed a part of Web-based news publications, whether major (e.g. the San Francisco Chronicle and CNN Interactive) or minor (e.g. university and high school newspapers).
- (b) Discussion group pages: there were 433 pages that were largely from the archives of internet discussion groups.
- (c) Humour pages: I collected 108 pages devoted to jokes, parodies, and other non-serious treatments of Ebonics. Search engines located many others, but since their descriptions often suggested they were likely to be repetitions of pages I had already viewed, I was not as diligent about checking and cataloguing this type of page as I was with some other types, so they may well be under-represented in my catalogue.

Over 40% of the pages were difficult to subclassify and were thus essentially placed in a category of 'leftovers':

- (d) General pages, which included 635 pages of such things as scholarly or political discussion; organisational or personal opinion; course materials, projects, and essays; and pages of relevant links. Some of the sites included pages (or copies of pages) of the other three types in addition to such material.

An interesting educational problem posed by Web-based materials is that there may be little in the appearance of the pages themselves to distinguish clever crackpots from scholarly opinion. Users of the Web probably need to develop better critical abilities than they might need for material printed under the auspices of reputable publishers and scholarly societies.

2. How I Found It: Data Gathering Techniques

There are actually various ways to locate relevant Web pages. A popular place to start is with one of the various search facilities available, and pages found in this way may also have relevant links that can be followed up.

Ten different search engines were ultimately employed to locate pages relating to Ebonics. Since some of these search engines would list as many as a thousand pages or more, not all the pages were actually viewed and catalogued. Instead the results of the search engines were sampled, though not highly systematically: generally more attention was given to those pages listed earlier by search engines, and also to the ones listed later that tended to stand out for some reason, e.g. because their description suggested they were especially relevant or perhaps just different in nature from other pages that had been found.

Once a page had been chosen from the results of a search, however, all of its relevant links were followed up. This sometimes led to dozens of additional pages. At an extreme, there was one page whose links ultimately led to about 250 different pages. Naturally there was quite a lot of overlapping of the links given on various pages, so as soon as a link took me to a page I had found earlier, I did not need to follow further links a second time. The patterns of linking were also interesting for how they divided the Ebonics pages into groups, like smaller or larger “islands” of knowledge. Naturally later pages tended to have more inclusive sets of links, although since earlier pages were sometimes revised the links could go in both directions.

There are other ways to locate relevant Web pages in addition to the above. A variation on the above is to search not for the topic itself, in this case ‘Ebonics’, but on something or someone else involved in the matter. Thus a few searches were run on the names of such people as Toni Cook, one of the people involved in Oakland School Board decision. This particular search yielded some useful pages that were less obvious from the search on ‘Ebonics’, such as a key interview with this person. At the same time such searches may also yield a great number of irrelevant pages, in this case about half a dozen other people with the same name. One of the advantages of searching for a topic like ‘Ebonics’ is that this word has only the one meaning and was not in fact widely used before the Oakland decision.

Searching on such related terms as ‘Toni Cook’ is an example of what might vaguely be called “looking in likely places”. As a somewhat different example of this, having found a number of Ebonics Web pages on commercial sites run by AOL, Geocities, and Tripod, I decided to search further within those sites using their own internal search facilities. This located large numbers of sites that were often not prioritised very highly by more general search engines and thus not always listed by them (most search engines actually list only a few hundred to about a thousand pages, even though they may have many more “hits” than this). By the same token, of course, these additional pages were not necessarily ones a searcher would find particularly important; many seemed to be repetitions or variants of Ebonics “humour” pages, for example.

There are also tricks one can do with site URLs (or “site addresses”). For example, a search engine would sometimes lead me to a page that seemed likely to be part of a larger collection, although the page itself might not have any links to related pages. For example, one search engine found certain pages which had only the title ‘Ebonics’ and a short discussion, with no links and no information about authorship. What one can do in such cases is to try successively shortened versions of the same URL. In this particular case one of the URLs was the first of the following, which can be successively shortened as shown in the subsequent lines:

`http://mbhs.bergtraum.k12.ny.us/user/b5616/opin04.html`

`http://mbhs.bergtraum.k12.ny.us/user/b5616/`

then `http://mbhs.bergtraum.k12.ny.us/user/`

then <http://mbhs.bergtraum.k12.ny.us/>

Not all of the shortened URLs will work — some simply produce error messages — but in this particular case the last of these turned out to be the homepage of Murray Bergtraum High School in New York City, and by following links from that page I found that the original pages was just one of fifty pieces of student work on Ebonics within a course on ‘Cyber English’.

3. How much is out there?

There must now be hundreds of millions of pages available on the World Wide Web: in early 1999 the Northern Light search engine claimed to cover 120 million Web pages. How many of these pages actually relate to Ebonics?

Table 1 shows the number of “hits” for ten search engines used to search for the term ‘Ebonics’ in about July 1997 and/or in January 1999. Two numbers are given in each case: the second, generally larger, number following the slash (/) is the total number of “hits” claimed by the search engine, but only the first, generally smaller, number of pages in front of the slash were actually listed. In January 1999, for example, Altavista claimed to have 24,738 “hits” for the word ‘Ebonics’ — the highest number of hits noted — but it would only list the first 200 of these.

Table 1. Numbers of “hits” for ‘Ebonics’

<u>search engine</u>	<u>July 1997</u>	<u>January 1999</u>
Altavista	200/~5000	200/24,738
Excite	1010/3481	1020/2078
HotBot	1100/3239	1000/4460
Inference Find!	130/130	51/51
Infoseek	350/378	304/3454
Looksmart	200/5540	200/24,720
Lycos	464/464	<1000/??
Northern Light	(untried)	>1150/9184
Webcrawler	182/182	169/169
Yahoo!	49/49	35/35

Were there really as many as 24,738 Web pages on Ebonics in January 1999? If so, why did no search engine other than Altavista and Looksmart (which appear to draw on the same database) find more than 9184 (in the case of Northern Light)?

By somewhat convoluted reasoning Black (1999d, sec. 5) estimated that they may actually have been as few as five to seven thousand pages on Ebonics freely available on the Web in January 1999. In part the reasoning has to do with evidence that the figures given by Altavista and Looksmart may include many pages that don’t actually mention Ebonics; at least one of these engines gave a ‘word count’ for the term ‘Ebonics’ that was actually less than the total number of hits. In addition, while Northern Light claimed 9184 hits, these include some material available for a fee from a database rather than freely available on the Web. In addition, some Web pages have disappeared even while they are still listed by search engines, as discussed in the following section.

4. How Long Does It Stay?

Once you locate a relevant Web page, what are the chances that it’s going to stay put? Is it safe to cite Web pages in the way we cite books and articles, expecting them to be there when someone else wants to consult them?

The answer is ‘No, not really.’ Firstly, for some of the searches I did in July 1997 I systematically checked and recorded all of the top twenty to seventy items. Even at that time about ten percent of these Web pages had become unobtainable, at least at the given URLs, even though most would have been created within the preceding seven months. By January 1999, a year and a half later, an additional 33% of the items located in 1997 could no longer be found, making a total of 43%.

Table 2. 1997 search results no longer on the Web

<u>search engine results</u>	<u>in July 1997</u>	<i>in January 1999</i>
Excite top 20	5%	30%
Infoseek top 70	13%	69%
Lycos top 50	6%	16%
Yahoo! all 56	10%	41%
All four engines: top 196	10%	43%

Different types of pages do not disappear at the same rate. From Table 3 below you can see the ‘humour’ pages are especially ephemeral, with only a fifth remaining two years after the Oakland policy announcement. The messages within discussion groups are also relatively unstable; perhaps those managing the discussion groups tend to have the good sense not to archive our sometimes incautious patter forever. That as many as a third of the ‘news’ pages disappear may seem surprising, but to some extent many are simply removed from freely accessible areas of the Web but can still be located and obtained from news archives for a fee. (A fee such as US \$2.95 for a brief article may, of course, be greater than the original cost of the whole newspaper.)

Table 3. 1997 search results no longer on the Web in 1999, by topic

General (270 pages)	21%
News (158 pages)	33%
Discussion (271 pages)	66%
Humour (29 pages)	79%
All 728 pages	43%

5. How Good Are The Search Engines?

Anyone who has used a search engine on the Web has probably had the experience of being directed to sites that were not particularly useful, and occasionally perhaps of no use whatsoever. As a striking example, eight of the top ten ‘Ebonics’ sites found by the search engine HotBot in 1998-99 were actually sex sites which apparently had used such terms as ‘Ebonics’ to try to attract search engines. Interestingly, other search engines did not seem to be taken in by these sites, certainly not enough to list them among the top ten.

From a study of all sites sampled it was easy to identify the most important sites in terms of the extent to which they deal with Ebonics and (especially) how many links they have to related sites. It was then possible to see to what extent search engines had located more relevant and important sites. One way in which this was done was by ranking the first twenty sites located by each search engine on a scale of 0 to 6 as follows.

0. Has no mention of Ebonics whatsoever.
1. Mentions Ebonics but not provide any useful information on it.
2. A bit better than 1 above but not as useful as 3 below.
3. Provides minimally useful example or commentary, e.g. a relevant paragraph a joke, or perhaps a couple of relevant links.
4. Provides more substantial discussion and/or up to ten relevant links.
5. Provide an especially in-depth treatment over a number of Web pages and/or by citing some ten to twenty relevant links.
6. Has over twenty links to significant pages.

For eight search engines used in 1998-99 an average ranking was calculated for the first twenty hits, as shown in Table 4. With ‘Ebonics’ as the search term, the average ranking ranged from 2.2 to 3.5, with only three search engines listed sites averaging more than the ‘minimally useful’ ranking of 3. Thus, despite the distinctiveness of the term ‘Ebonics’, the majority of pages listed in the top twenty tended to be of relatively little value. Most search engines did noticeably better — mostly from 3.4 to 4.3 — when searching for documents that contained both ‘Ebonics’ and ‘Oakland’. Interestingly the two search engines that improved least were those that scored worst and best on the search for ‘Ebonics’ alone: Webcrawler’s 2.2 improved only to 2.5, while Northern Light’s 3.5 remained unchanged.

Table 4. Average importance of top 20 ‘hits’ (on a scale of 0 to 6)

<u>search engine</u>	<u>'ebonics'</u>	<u>+ebonics+oakland'</u>
Excite	3.4	4.1
HotBot	2.8	4.3
Inference Find!	2.3	3.4
Infoseek	2.3	4.2
Looksmart	2.7	3.8
Lycos	3.0	3.9
Northern Light	3.5	3.5
Webcrawler	2.2	2.5

(Only eight of ten search engines were included in the above because Altavista generally produced the same results as Looksmart, while Yahoo! does not rank sites in the way other search engines do. Yahoo! lists relatively small numbers of what are clearly relevant sites (i.e. those ranking 3 or higher) and thus would tend to score relatively high.)

Another measure of relative effectiveness was to check and see how many of the most important (i.e. level 5 and 6) sites available at the time were actually listed at all by the search engines used, i.e. even listed among the top thousand if the search engine listed that many sites. For those search engines used in 1997 it was found that none listed more than eight of the eighteen most important sites on the Web at that time, and one of those eighteen sites was not listed at all by a search engine (it was found by following links instead). A year and a half later most search engines tended to perform just slightly better. Clearly you can't expect any one search engine to lead you to all of the most important sites.

6. Conclusion

Although you won't yet find everything on the World Wide Web, it can certainly be a valuable source of popular and scholarly opinion on such matters as language and education. The amount of relevant information on the Web is often enormous, but some of it also tends to be ephemeral, and few of the search engines used in this study in 1997-99 were able to lead one very quickly to the most valuable sources. Mining the Web for data seemed rather like working a low grade gold mine: accumulating precious treasure required a great deal of work or luck, at least with the somewhat blunt search tools available at the time.

References:

- Black, Paul 1999a, Bilingual education in the NT, electronic document available from <<http://www.ntu.edu.au/education/csle/issues/ntbiling.html>>.
- Black, Paul 1999b, The Ebonics Issue, electronic document available from <<http://www.ntu.edu.au/education/csle/issues/ebonics.html>>.
- Black, Paul 1999c, Proposition 227, electronic document available from <<http://www.ntu.edu.au/education/csle/issues/prop227.html>>.
- Black, Paul 1999d, World Wide Web coverage of the Ebonics issue, electronic document available from <<http://www.ntu.edu.au/education/csle/research/ebonics/eb0.html>>.
- Center for Applied Linguistics 1997, Ebonics information page, electronic document obtained 28 Dec. 1998 from <<http://www.cal.org/Ebonics/>>.
- [Linguistic Society of America] [1997] LSA Resolution on the Oakland "Ebonics" issue, electronic document obtained 26 Dec. 1998 from <<http://linguistlist.org/topics/ebonics/lsa-ebonics.html>>.
- Williams, Robert L. 1975, Ebonics: The True Language of Black Folks, Institute of Black Studies, St Louis.