



Disability Rights Promotion International

Developing a Global Methodology for Monitoring Human Rights in Media Depictions of Disability

The DRPI NN team at UB spring/fall 2008:

Brenda Battleson

Hao Chen

Carolyn Evans

Alaina Iacobucci

Anne Solbu-Slowe

Joseph Woelfel

Ezra Zubrow

1. Description of methods used to monitor media coverage of disability

a. Neural networks analysis

Perhaps the largest body of data available to the social scientist is text. The analysis of text, however, is particularly difficult. Two generic methods generally define text analysis.

The first type of text analysis involves careful study of text by skilled interpreters. Discourse and content analysis are generally assumed to produce the deepest and most detailed understanding of text, but is by their labor-intensive nature limited to relatively small texts. Although software to assist content analysis exists, it is mainly to aid the analyst in her/his analysis, and does no analysis itself. Although such software can indeed be extensive and useful, it is mainly a way to search for concepts, phrases, keywords in context (KWIK), and other linguistic structures that must be provided in advance by human analysts who know what they are looking for.

Proximity analysis

A second type of computer software, however, actually performs analysis with little or no guidance by human analysts. These algorithms generally work by calculating the proximities of words in the text. The simplest of these, and the most common, utilize a simple co-occurrence algorithm. In the co-occurrence model, the text is divided into discrete parts (called “cases” in the Catpac model). These cases could be linguistic units, such as sentences or paragraphs, or arbitrary units, such as pages, chapters, articles, or even arbitrary chunks. Co-occurrence software counts the number of times each pair of words co-occurs in the cases, and these numbers of co-occurrences are used as a measure of the similarity of the words – the higher the co-occurrence, the more similar the words are judged to be.

More powerful proximity models add a “sliding window” instead of the discrete parts or cases. In such models a “window” of several words “slides” through the text. With a three-word window, for example (the size most commonly used by Danowski, 2008) the program will begin with the first three words of a text in the window. Each of these three words will be counted as co-occurring, since they are in the window simultaneously. The window will then slide one word, and words 2, 3 and 4 will now co-occur in the window. When the window has slid through the entire text, the result is a word-by-word matrix of similarities based on co-occurrences in the sliding window.

Neural text analysis

The most powerful model – the one employed in Catpac and Katmandu (Chen, et. al., 2008) adds an artificial neural network (Woelfel, 1993) to the proximity system. Catpac and its larger sibling Katmandu can utilize either a simple case based or window based

co-occurrence model, but differ fundamentally from other proximity models. Catpac does not count co-occurrences. Rather, each unique word¹ in the text is assigned an artificial neuron. A neuron, however complicated in actual organic nervous systems, can be considered a simple unit which can be dormant or active. In Catpac, the level of activity for any neuron can vary continuously between zero (dormant) and 1.1.

As in an organic system, these neurons may or may not be interconnected in a way that allows them to share their activation. Thus, if two neurons are interconnected, when one becomes active, it can communicate its activation to the other. If the connection between the two is strong, a large portion of the activation can be communicated; if the connection is weak, proportionally less activation can be communicated. The system of interconnected neurons is referred to as a “neural network.”

The operation of the network

When any neuron is in a window (or a case) its activation is set to 1.0. Catpac’s window size can be set by the user, but its default value is 7, so, in general, seven words will be in the window at all times. When the program begins, these will, of course, be the first seven words of the text, so the activation of each of the first seven words will be set to 1.0.

In an actual organic brain, when some neurons are activated, they will communicate their activation to other neurons to which they are connected. But when Catpac is reading the first seven words of the text, none of the neurons are yet connected and *only* the first seven words will be active. The program then polls all neurons, determines their levels of activation, and increments the connection between all pairs of active neurons proportionally to their activations. Obviously, the first seven words in the window – and *only* those words -- will all be active, and the connections among them will all be incremented.

The window then slides one word to the right, so words 2 through 8 are now in the window, and the process is repeated. As the window continues to move through the text in this manner, it is likely that some of the words in the window may be connected to other words not in the window. Thus, activation of the words in the window may lead to the activations of other words as well. Again, the connection among all active words is strengthened proportionally to their co-activation.

At each cycle all connections are weakened by a small amount to simulate a constant rate of forgetting, which is natural to all organisms. Due to the forgetting, only those words that are frequently co-active will be tightly connected. Casual connections due to chance or other minor factors are forgotten. Also, at each cycle, the matrix of connection strengths is centered by subtracting the mean connection strength from each cell and normalized so that the largest connection strength is 1.0.

¹ Multiple occurrences of the same word are considered one “unique” word.

As with the other propinquity based methods, the result of Catpac analysis is a word-by-word matrix. But instead of counts of co-occurrences in the text's cases or moving windows, Catpac's matrix contains connection strengths among the neurons and thereby, the words they represent; these connections may vary in strength and are not just on or off, connected or not. Accordingly, the patterns of connections among the neurons in the network are simulacra of the patterns the network encountered in the text. As in an organic brain, connections among the neurons form the network's memory of what it has encountered. Also, as in an organic brain, the memories represent only the systematic patterns of experience; for the most part, minor or infrequent relationships have been forgotten. The connections strengths, then, represent the core concepts in the text.

Data presentation

Whatever method is used to obtain the square words-by-words matrix, the matrix needs to be represented in a simple and graphic way. The most common procedure is cluster analysis, in which the matrix is analyzed to find out how the words group together or "cluster." Catpac provides the option for 7 different cluster analysis algorithms; the default algorithm is Ward's Method. The output of the cluster analysis is arrayed in a dendrogram (sometimes spelled "dendrogram"), which is illustrated in Table 1.

Words are written vertically across the top of the dendrogram. Underneath the words, patterns of carets underlie the words that go together. The higher the structure of the carets, the more closely the words above it are connected. At the extreme right of the dendrogram, the highest part of the structure occurs under the words "people" and "disabilities", indicating that these are the two most closely connected words in the text. To the left, a slightly lower structure underlies the words "special" and "education." At the highest level, "people" and "disabilities" form one cluster, and "special" and "education" form another.

WARDS METHOD

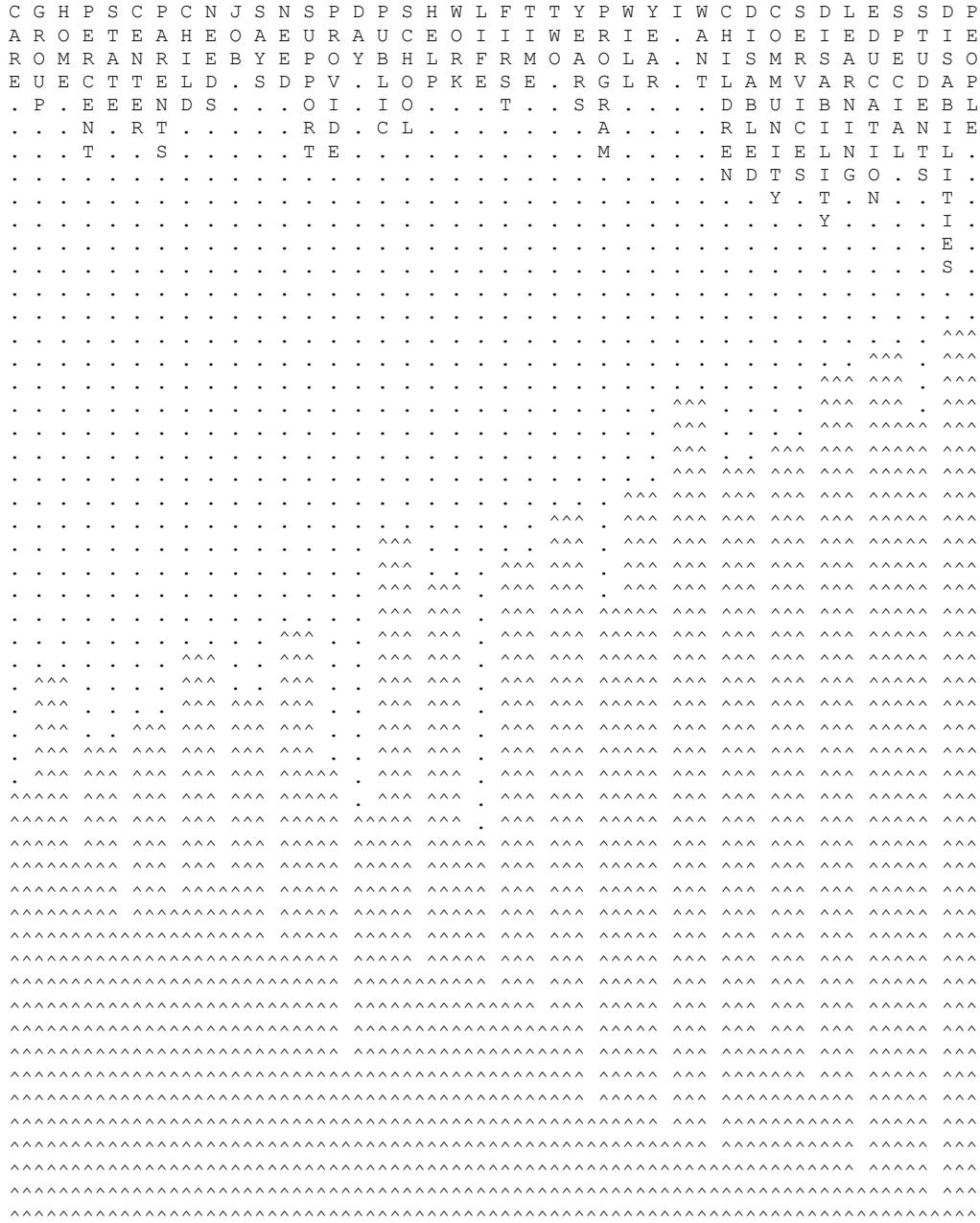


Table 1: Dendrogram for the term “disabilities”

To the left and slightly lower, “learning” and “disability” form a third cluster. Slightly lower and to the right, we can see that “students” joins the “special education” cluster. As we continue to move downward in the dendrogram, we can see new clusters forming, and previously distinct clusters merging into more complex clusters. To the left of the dendrogram, a large cluster contains the terms “provide”, “support”, “state”, “child”, “needs”, “parent”, “home,” “care”, and “center”. To the right of this cluster there is a “public school” cluster, including also “first”, “time”, “help”, “life” and “work”. To the

right of this is a cluster including “disabled”, “children”, “community”, “services”, “learning”, and “disabilities.” To the right of this is a cluster concerning “special”, “education”, and “students.” Last, there is the very strong cluster, “people” and “disabilities.” These clusters represent the most important concepts which underlie the text.

A second, more precise method of presenting the same information is the perceptual map. Since the connections among the words can be interpreted as “distances,” it is possible to calculate the coordinates of the words in a visual map. Catpac uses the Galileotm algorithm (Woelfel & Fink, 1980) to compute these coordinates, which are displayed by the program ThoughtView (Woelfel, 1993) in Figure 1.

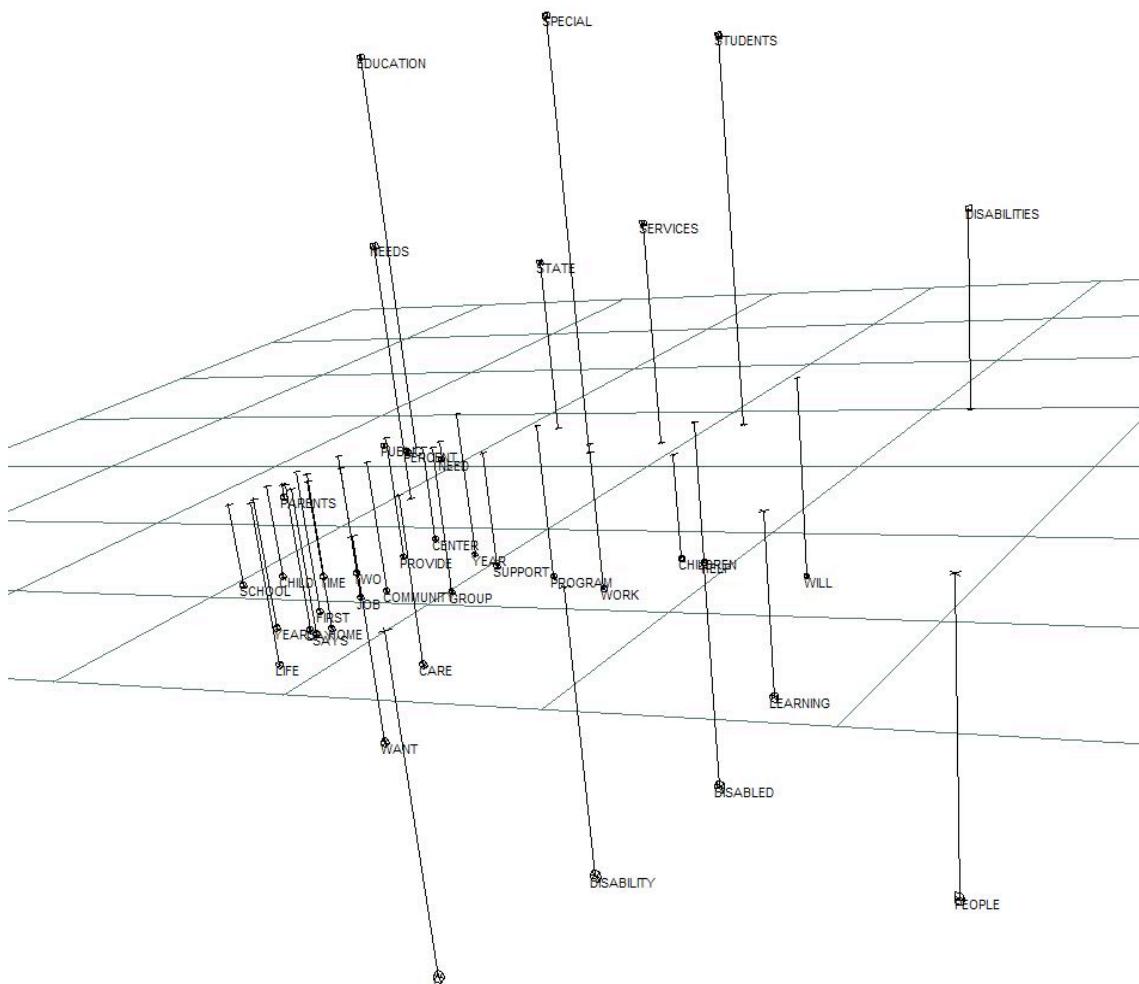


Figure 1. ThoughtView Plot of Disabilities

The “People” and “disabilities” cluster is shown by these two terms standing off to the right of the picture. The “school” cluster is tightly grouped at the left side of the picture, while the “special education” cluster is at the top. To the right of the “school” cluster the “learning”, “disabled”, “children”, “community”, “services” cluster can be seen to the right of the “school” cluster.

ThoughtView allows real time interaction with the picture – it can be rotated, zoomed, panned and otherwise manipulated to help understand the structure. It is also possible to magnify the plot using the zoom feature in a typical word processor. By doing this, and referring back to the dendrogram, it is possible to get a good visual understanding of how the concepts relate to each other.

Statistics

Catpac also presents extensive quantitative information to aid in analysis, most importantly the frequencies of occurrences² of the words. Table 2 shows that the “disabilities” text contained 9,711 words, followed by lists of the top 40 words in order of frequency of occurrence and in alphabetical order. Catpac allows the user to specify how many unique words to analyze; in this analysis the maximum number of words to analyze was set at 40. The frequency of occurrences for each unique word is listed as both a count and as a percentage of all word occurrences. (“Case freq” and “case pct” refer to the percentages of all cases in which the words occurred if casewise analysis was used; in this study only the sliding windows method was employed.) Thus the most frequently occurring word is “disabilities”, which occurred 1063 times and represented 10.9% of all occurrences. The least frequently occurring words of the top 40 words were “child”, “day”, and “want”, each of which occurred 113 times, for 1.2% of all occurrences.

² Catpac offers the option of excluding certain words from analysis. These are generally articles, prepositions and other words that experience has indicated have no significance, yet occur frequently enough to distort analysis. Often, for example, the most frequently occurring word in a text is “the”. A default list of excluded terms is included in a file called default.exc, but this file may be edited or substituted by the user.

TOTAL WORDS	9711	THRESHOLD	0.000
TOTAL UNIQUE WORDS	40	RESTORING FORCE	0.100
TOTAL EPISODES	123938	CYCLES	1
TOTAL LINES	7889	FUNCTION	Sigmoid (-1 - +1)
		CLAMPING	Yes

DESCENDING FREQUENCY LIST					ALPHABETICALLY SORTED LIST				
WORD	FREQ	PCNT	CASE FREQ	CASE PCNT	WORD	FREQ	PCNT	CASE FREQ	CASE PCNT
DISABILITIES	1063	10.9	7398	6.0	CARE	116	1.2	808	0.7
PEOPLE	919	9.5	6378	5.1	CENTER	143	1.5	1001	0.8
I	568	5.8	3808	3.1	CHILD	113	1.2	781	0.6
WILL	454	4.7	3157	2.5	CHILDREN	266	2.7	1856	1.5
DISABILITY	366	3.8	2529	2.0	COMMUNITY	189	1.9	1321	1.1
DISABLED	335	3.4	2323	1.9	DAY	113	1.2	768	0.6
SCHOOL	323	3.3	2223	1.8	DISABILITIES	1063	10.9	7398	6.0
SERVICES	308	3.2	2143	1.7	DISABILITY	366	3.8	2529	2.0
STUDENTS	298	3.1	2067	1.7	DISABLED	335	3.4	2323	1.9
CHILDREN	266	2.7	1856	1.5	EDUCATION	244	2.5	1685	1.4
YEAR	259	2.7	1785	1.4	FIRST	141	1.5	987	0.8
WORK	249	2.6	1729	1.4	GROUP	173	1.8	1201	1.0
EDUCATION	244	2.5	1685	1.4	HELP	217	2.2	1512	1.2
SPECIAL	242	2.5	1684	1.4	HOME	125	1.3	871	0.7
STATE	232	2.4	1614	1.3	I	568	5.8	3808	3.1
HELP	217	2.2	1512	1.2	JOB	116	1.2	803	0.6
YEARS	215	2.2	1493	1.2	LEARNING	172	1.8	1197	1.0
PROGRAM	203	2.1	1415	1.1	LIFE	134	1.4	927	0.7
COMMUNITY	189	1.9	1321	1.1	NEED	139	1.4	971	0.8
GROUP	173	1.8	1201	1.0	NEEDS	145	1.5	1015	0.8
LEARNING	172	1.8	1197	1.0	PARENTS	123	1.3	850	0.7
TIME	159	1.6	1113	0.9	PEOPLE	919	9.5	6378	5.1
NEEDS	145	1.5	1015	0.8	PERCENT	117	1.2	755	0.6
CENTER	143	1.5	1001	0.8	PROGRAM	203	2.1	1415	1.1
FIRST	141	1.5	987	0.8	PROVIDE	115	1.2	805	0.6
NEED	139	1.4	971	0.8	PUBLIC	133	1.4	931	0.8
SUPPORT	138	1.4	965	0.8	SAYS	119	1.2	825	0.7
LIFE	134	1.4	927	0.7	SCHOOL	323	3.3	2223	1.8
PUBLIC	133	1.4	931	0.8	SERVICES	308	3.2	2143	1.7
HOME	125	1.3	871	0.7	SPECIAL	242	2.5	1684	1.4
PARENTS	123	1.3	850	0.7	STATE	232	2.4	1614	1.3
SAYS	119	1.2	825	0.7	STUDENTS	298	3.1	2067	1.7
PERCENT	117	1.2	755	0.6	SUPPORT	138	1.4	965	0.8
CARE	116	1.2	808	0.7	TIME	159	1.6	1113	0.9
JOB	116	1.2	803	0.6	TWO	114	1.2	787	0.6
PROVIDE	115	1.2	805	0.6	WANT	113	1.2	786	0.6
TWO	114	1.2	787	0.6	WILL	454	4.7	3157	2.5
CHILD	113	1.2	781	0.6	WORK	249	2.6	1729	1.4
DAY	113	1.2	768	0.6	YEAR	259	2.7	1785	1.4
WANT	113	1.2	786	0.6	YEARS	215	2.2	1493	1.2

Table 2: Frequency counts of words for Disabilities

The greatest virtue of neural network analysis of text is that it gives the deepest level of analysis that can be achieved automatically, that is, without the intervention of human analysts. This makes it ideal for use in scanning very large bodies of text early, as a screening device to detect areas of interest that might warrant deeper scrutiny via discourse or content analysis.

This capability to provide relatively deep analyses of very large amounts of text without expert intervention suggests two effective ways to implement the technology. First, a

very small non-technical staff would be sufficient to monitor worldwide flows of disability rights information and post results to a website on an ongoing basis. Participating researchers could then access these data from anywhere in the world and quickly be apprised of emerging trends. Second, the technology is suitable for use by non-technical personnel whose expertise lies in areas other than computing and text analysis. This means that stand-alone neural analysis kits could be provided to participating researchers around the world to allow local analysis of media on demand.

These references come from the draft for all sections of *Developing a Global Methodology for Monitoring Human Rights in Media Depictions of Disability*:

References

- Atkin, C., et al. (2008). A Comprehensive Analysis of Breast Cancer News Coverage in Leading Media Outlets Focusing on Environmental Risks and Prevention. *Journal of Health Communication*, Vol. 13 Issue 1, p3-19.
- Bernt, J.P. and Greenwald, M.S. (1993). *Coverage of gays, lesbians, bisexuals and the HIV/AIDS: A content analysis of seventeen metropolitan daily newspapers*. Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication, Kansas City, Mo.
- Berry, T.R., Wharf-Higgins, J. & Naylor, P.J. (2007). SARS Wars: An Examination of the Quantity and Construction of Health Information in the News Media. *Health Communication*, Vol. 21 Issue 1, p35-44.
- Clogston, J.S. (1993a, August). *Changes in coverage patterns of disability issues in three major American newspapers, 1976-1991*. Paper presented at the Annual Meeting of the Association for the Education in Journalism and Mass Communication, Kansas City, Mo.
- Chen, H., Evans, C., Battleson, B., Zubrow, E., & Woelfel, J. (2008, January 27). Procedures for the Precise Analysis of Very Large Textual Datasets. Paper presented at the Sunbelt XXVIII, INSNA Social Networking Conference, St. Pete Beach, Florida.
- Clogston, J.S. (1993b, March 8). Media models. Personal communication.
- Clogston, J.S. (1992a, August). *Fifty years of disability coverage in the New York Times, 1941-1991*. Paper presented at the Annual Meeting of the Association for the Education in Journalism and Mass Communication, Montreal, Canada.
- Clogston, J.S. (1992b, August). *Coverage of persons with disabilities in prestige and high circulation dailies*. Paper presented at the Annual Meeting of the Association for the Education in Journalism and Mass Communication, Montreal, Canada.
- Clogston, J.S. (1992c, March). *Journalists' attitudes toward persons with disabilities: A survey of reporters at prestige and high circulation dailies*. Paper presented at the spring conference on Women, Minorities, and the Mass Media, Association for the Education in Journalism and Mass Communication, Atlanta, Ga.
- Clogston, J.S. (1991). *Reporters' attitudes toward and newspaper coverage of persons with disabilities*. Unpublished doctoral dissertation at Michigan State University.
- Clogston, J.S. (1990a). *Disability Coverage in 16 Newspapers*. Louisville: Advocado Press.

Clogston, J.S. (1990b, June). *Perceptions of disability in the New York Times*. Paper presented at the annual meeting of the Society for Disability Studies, Washington, D.C.

Clogston, J.S. (1989, August). *A theoretical framework for studying media portrayal of persons with disabilities*. Paper presented at the Annual Meeting of the Association for the Education in Journalism and Mass Communication.

Danowski, J.A. (2008). WordLink, version Infinity. [Computer program]. Chicago: University of Illinois at Chicago.

Davidson, A. E. & Wallack, L. (2004). A Content Analysis of Sexually Transmitted Diseases in the Print News Media. *Journal of Health Communication*, Vol. 9 Issue 2, pp. 111-117.

Fairclough, N. (1995). *Media discourse*. London: Edward Arnold.

Fairclough, N. (2000). *Language and power* (2nd ed.). New York: Longman.

Gitlin, T. (1980). *The whole world is watching*. Berkeley, Ca.: University of California Press.

Graber, D. (1989). Content and meaning. *American Behavioral Scientist*, 33:2, 144-152.

Haller, B., Dorries, B. & Rahn, J. (2006, January). Media labeling versus the U.S. disability community identity: A study of shifting cultural language. *Disability & Society*, January 2006, Vol. 21, No. 1.

Haller, B. & Ralph, S. (2001). Content analysis methodology for studying news and disability: Case studies from the United States and England. *Research in Social Science and Disability*. JAI Press, Vol. 2.

Haller, B. (1995). *Disability Rights on the Public Agenda: News Media Coverage of the Americans with Disabilities Act*. Unpublished doctoral dissertation, Temple University, Philadelphia, Pa.

Harlan, C. (1996, April 17). Hispanic roles on TV still few, group finds. *Austin American-Statesman*, Sec. A.

Hayes, M. et al. (2007). Telling stories: News media, health literacy and public policy in Canada. *Social Science & Medicine*, Vol. 64 Issue 9, p1842-1852.

Henry, F., & Tator, C. (2002). *Discourses of domination*. Toronto: University of Toronto Press.

Higgins, P.C. (1992). *Making disability: Exploring the social transformation of human variation*. Springfield: Ill.: Charles C. Thomas Publisher.

- Janowitz, M. (1968). Harold Lasswell's contribution to content analysis. *Public Opinion Quarterly*, (Winter), p. 648.
- Jorgensen, P. (2002). Continuous analysis of internet text by artificial neural network (Doctoral dissertation, University at Buffalo, 2002).
- Krossel, M. (1988). Handicapped heroes and the knee-jerk press. *Columbia Journalism Review*, 46-47.
- Louis Harris and Associates, Inc. (1991). *Public attitudes toward people with disabilities*. New York: National Organization on Disability.
- McCombs, M. (1992). Explorers and surveyors: Expanding strategies for agenda-setting research. *Journalism Quarterly*, 69:4, pp. 813-824.
- McCombs, M. & Shaw, D. (1993). The evolution of agenda-setting research. *Journal of Communication*, 43:2, pp. 58-67.
- McCombs, M. & Shaw, D. (1972). The agenda-setting function of the press. *Public Opinion Quarterly*, 36, 176-187.
- McGregor, Sue L. T. (2003a). *Critical Discourse Analysis--A Primer*.
<http://www.kon.org/archives/forum/15-1/mcgregorcda.html>. Accessed December 19, 2008.
- McGregor, Sue L.T. (2003b). *Critical Science: A Primer*.
<http://www.kon.org/archives/forum/15-1/mcgregorcs.html>. Accessed December 19, 2008.
- McQuail, D. (1989). *Mass Communication Theory*. London: Sage.
- Mitchell, L. R. (1989). Beyond the Supercrip syndrome. *Quill*, pp. 18-23.
- OECD (2003) *Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People*.
- Poindexter, P. M. & McCombs, M. E. (2000). *Research in Mass Communication: A practical guide*. Boston: Bedford/St. Martin's.
- Raman, P. et al. (2008). Portrayals of Older Adults in U.S. and Indian Magazine Advertisements: A Cross-Cultural Comparison. *Howard Journal of Communications*; Vol. 19 Issue 3, pp. 221-240.
- Ryan, M. & Owen, D. (1976). A content analysis of metropolitan newspaper coverage of social issues. *Journalism Quarterly*, 53, 634-640, 671.

Scotch, R.K. (1988). Disability as the basis for a social movement: Advocacy and politics of definition. *Journal of social issues*, 44:1, pp. 159-172.

Shoemaker, P. J. & Reese, S. D. (1996). *Mediating the Message, Theories of influences on mass media content*. NY: Longman.

Slater, M. D. et al. (2008). News Coverage of Cancer in the United States: A National Sample of Newspapers, Television, and Magazines. *Journal of Health Communication*, Vol. 13 Issue 6, pp. 523-537.

Taylor, W.L. (1965). Gauging mental health content in the mass media. *Journalism Quarterly*, 34, 191-201.

Taylor-Clark, K. A. et al. (2007). News of disparity: Content analysis of news coverage of African American healthcare inequalities in the USA, 1994–2004. *Social Science & Medicine*, Vol. 65 Issue 3, pp. 405-417.

UN Web Services Section, (2006), “Some Facts about Persons with Disabilities”, United Nations, Department of Public Information. Available at:
<http://www.un.org/disabilities/convention/facts.shtml> (accessed November, 2008).

UN Development Programme, quoted in UN Web Services Section, *Ibid*.

van Dijk, T. A. (1985). “Structures of news in the press,” in van Dijk, (Ed.) *Discourse and Communicatio*. Berlin: De Gruyter, pp. 69-93.

van Dijk, T. A. (1988). *News as discourse*. Hillside, NJ: Erlbaum.

van Dijk, T.A. (2001). ”Critical Discourse Analysis,” In D. Tannen, D. Schiffrin & H. Hamilton (Eds.), *Handbook of Discourse Analysis*. Oxford: Blackwell, pp.352-371.

Voakes, P. S., Kapfer, J., Kurpius, D. & Shano-yeon Chern, D. (1996). Diversity in the news: A conceptual and methodological framework. *Journalism and Mass Communication Quarterly*, pp.582-593.

Woelfel, J. (1992). *Communication and science*. New York: McGraw-Hill.

Woelfel, J. (1993). Cognitive processes and communication networks: A general theory. In W. Richards & G. Barnett (Eds.) *Progress in Communication Sciences 12*. (pp. 21-42). Norwood, N.J.: Ablex Pub. Corp.

Woelfel, J., & Fink, E. (1980). *The measurement of communication processes: Galileo theory and method*. New York: Academic Press.

Woelfel, J., & Richards, W. (1989, November 12). A general theory of intelligent, self referencing networks. Paper presented at the Buffalo Conference on Networks, State University of New York at Buffalo, Buffalo, New York.

Woelfel, J., & Stoyanoff, N. J. (2007). The Galileo System: A rational alternative to the dominant paradigm for social science research. In *Freiberger Beiträge zur interkulturellen und Wirtschaftskommunikation: A Forum for General and Intercultural Business Communication*. Berlin: Peter Lang.

Zywica, J., Danowski J. (2008). The Faces of Facebookers: Investigating Social Enhancement and Social Compensation Hypotheses; Predicting Facebook™ and Offline Popularity from Sociability and Self-Esteem, and Mapping the Meanings of Popularity with Semantic Networks. *Journal of Computer-Mediated Communication*, 14(1), 1-34.