A COMPUTER CONTENT ANALYSIS APPROACH TO MEASURING SOCIAL DISTANCE IN RESIDENTIAL ORGANIZATIONS FOR OLDER PEOPLE

Donald G. McTavish
Department of Sociology
University of Minnesota
Minneapolis, MN 55455

Kenneth C. Litkowski
CL Research
20239 Lea Pond Place
Gaithersburg, MD 20879

Susan Schrader
Department of Sociology
Augustana College
Sioux Falls, SD 57105

Computer content analysis provides another approach to measuring aspects of social structure. Different social positions imply different social perspectives that are evident in language. A language-based measurement of distance between positions in an organization is described, using as data verbatim transcripts of interviews with occupants of positions in nursing homes talking about their organizational situation. Minnesota Contextual Content Analysis (MCCA), a computer content analysis approach, scores social perspectives in these texts and computes social distance as a function of differences between perspectives, facilitating an examination of social distance with other organizational and personal outcomes. Correlates of distance between roles across nursing homes suggest consequences for organizational structure and the meaning residents express about their experience. This content analysis permitted a relatively accurate identification of each respondent with a particular nursing home, a measurable aspect of organizational culture. The structure of these differences reveals important facets of organizational structure. The methodology used here is compared to techniques in information retrieval for characterizing documents by semantic vectors. This comparison suggests that MCCA captures finer-grained concepts. Semantic analysis of the MCCA categories using WordNet shows that they constitute semantic domains, whose further refinement may lead to better characterization of the identified differences.

1 Introduction

This paper describes a language-based approach to the measurement of the social distance between positions in an organization, as an example of using a content analysis technique that may usefully be extended to numerous other applications. In everyday interaction, differences in stylistic register and pragmatics are evident between managers and employees, doctors and patients, accountants and sales staff in *what* they commonly discuss as well as *how* these topics are presented. The approach discussed in this paper uses conversational interviews with occupants of positions in organizations in which they talk about their organizational situation. Computer content analysis

provides a practical tool for reliably coding social perspectives using these verbatim transcripts. Differences between perspectives is a function of social distance between statuses. This approach to the measurement of social distance in organizations facilitates an examination of the impact this feature of social structure upon other organizational and personal outcomes.

In this paper we describe a contextual content analysis approach to the measurement of cultural aspects of social distance between statuses in organizations. These distances are shown to be related to selected other features of organizations and individuals in the organizations. Conversational interview data from a study of administrators, staff and residents of nursing homes in Minnesota, USA are used to illustrate the approach.

2 Differences in Perspective Between Statuses

Several measures of social distance have been developed and various approaches have been used in analyzing them. (See, for example, Kadushin (1962), Kidwell & Booth (1977), McPhersion et al. (1987), Miller (1983), and Reiss (1961) for further details, as well as McTavish & Felt (1987) and McTavish & Schrader (1992) for consideration of these measures and approaches in the genesis of this study.) Here we focus on relevant concepts of social perspectives necessary to understand our results and to explicate our content analysis approach.

Qualified incumbents of different social positions learn expected role behavior and interact in somewhat different social circles by virtue of their positions; understand differing priorities, objectives, and obligations; and because of these differences, typically see their organizational situation in somewhat different ways. By virtue of their social position in a structure, they have somewhat different social perspectives.

The administrator of a nursing home, for example, is involved with residents and their kin, state regulators, volunteers, owners, staff and media, to name a few. A typical responsibility is to represent the organization to these various stakeholders. Residents, on the other hand, relate to other residents and to their own kin, to staff and administrators, and to various volunteers who may be around from time to time, primarily as the recipient of services. Different statuses promote different perspectives on the nursing home situation.

Particular statuses can be located in more general institutional structures. Thus we can identify a particular status such as "father" with a specific family organization as well as with a type of social institution (e.g. "the family"). More broadly, a status can be characterized in terms of the more general themes of institutional sectors. "Traditional" institutions (e.g. religion, family, legal institutions) emphasize normative standards for appropriate behavior and sanctions for deviance. Practical accomplishment of goals and success are themes emphasized in economic and production institutions. A given social position can be identified to a greater or lesser degree with the broader institutional sectors whose themes it most emphasizes. These institutional themes characterize a social position and are evident in the language used by occupants of a status. For example, while a minister might be expected to approach topics in a more traditional than pragmatic way, a church treasurer might typically emphasize pragmatic themes a bit more, stemming from the somewhat different perspectives involved in their different statuses in a church organization. The "same" types of status may have a somewhat different perspective in different social structures (societies, times, places). A salesperson for IBM has a different perspective than a salesperson for a car dealership and these have changed through time.

In this paper we measure social distance between statuses in terms of institutional perspectives, using the role behavior of participants' talking about their social position as an indicator of the institutional perspective of their social position. This is scored in a way that indicates the relative

emphasis on different institutional themes. Differences between scores in terms of these institutional perspectives is proposed as a measure of social distance.

3 A Language-based Measure of Social Distance.

Occupants of a social position are asked to talk in an open-ended fashion about what their context is like from their positional point of view. Open-ended conversations are a more relevant and direct solution to the measurement of differences in social perspective because they readily capture meanings and emphasis that the respondent wishes to express than many structured approaches. Ideas which the respondent introduces are conditioned by their experience in a particular status. This contributes to face validity by utilizing meanings a position-holder chooses to use. A verbatim transcript of these conversations is scored using a computer content analysis procedure to indicate similarity to themes characteristic of broader institutional perspectives. A social distance score is then computed. Computer scoring permits a reliable contrast of status-centered dialogue by avoiding coder reliability problems affected by a coder's own position, experience and fatigue.

4 The MCCA Approach.

The computer content analysis approach used here is called Minnesota Contextual Content Analysis (MCCA). MCCA is described in McTavish and Pirro (1990) and early work on a similar approach was reported by Cleveland, McTavish and Pirro (1974). Two kinds of normed scores are generated for each analyzed text. One kind of score shows the emphasis (called E-scores) placed on each of many idea categories. An idea category consists of a group of words which reflect a given idea or meaning. For example, the idea of "control" occurs when words such as "allow", "authorize", "insist", and "command" are used. The MCCA dictionary distinguishes 116 idea categories; words may be assigned to more than one category since they may have more than one sense; the assignments to categories were made judgmentally. Scores are "normed" against expected usage of the words in an idea category so that positive E-scores indicate an over-emphasis and negative E-scores indicate a relative omission of a given idea in the text. Normed scores are computed in a z-score-like fashion, contrasting category proportions with the expected probability of use of a given idea category, divided by a standard deviation of expected category usage across various social contexts. Expectations are based on the Kucera and Francis (1967) word counts and percentages.

The E-scores are computed for each of the categories and are the basis for the conceptual analysis. The pattern of connectedness of various ideas in a text is examined using a clustering routine. Similarity and distinction between texts in terms of emphasized patterns of ideas can be quantified as well. A distance between texts can be measured as a discrepancy between texts on their profile of relative use of the 117 categories (the 117th category is the "leftover" list of uncategorized words). The structure of conceptual differences shown in this proximity matrix can also be examined by clustering and other statistical techniques.

The second kind of score consists of a profile of four "social context" scores (C-scores). Organizational settings can be distinguished by the pattern of language used in these contexts, in part, because of different problems or tasks an organization typically encounters. Contextual content analysis utilizes these differential patterns of emphasis. Four vectors or contextual themes have been identified by factoring and experience with texts from different institutional sectors. Each context dimension is a function of the emphasis in the text across a large number of idea categories, and is represented in MCCA by a vector of weights; these marker contexts are experimental, empirically-

derived profiles of relative emphasis on each idea category. For example, although the idea of "deviance" can be used in any organization, it receives relatively greater emphasis in more traditional institutions. Together the profile of four social C-scores for a text measures how closely the text's pattern of language matches each of these four broad institutional themes. The reference themes are labeled as follows:

a) Traditional - a focus on norms and expectations for appropriate behavior. This is especially emphasized in text from judicial and religious organizations.

b) Practical - a focus on successful (efficient) goal accomplishment. This theme is most evident in business and work organizations.

c) Emotional - a focus on personal involvement, comfort, enjoyment, or leisure. This theme is typical of leisure or recreational organizations.

d) Analytic - a focus on objectivity, curiosity or interest. This emphasis is more pronounced in research and educational settings.

Usually there are no "pure types" and people describe their orientation to their setting in a "blend" or "profile" of emphases across the four contextual dimensions which, thus, locate the social context of the text. The social context of the text is used to disambiguate ambiguous words by accumulating contextual scores using weights reflecting the relative use of each conceptual category in the four social contexts. Table 1 illustrates E-scores and C-scores assigned to different texts, showing how different vectors are generated.

Where texts are generated by people occupying different status locations in an organization, the vector of four C-scores identifies the social location of their statuses. Distance between two C-score profiles can be represented by a standard euclidean distance computation. If the distance measure is zero, the two positions have the same profile of C-scores. The larger the distance measure, the greater the social context distance between the two positions. It becomes larger as they take on quite different contextual perspectives. If the measure between two organizational positions is very large, it is hypothesized that communication difficulties are likely to be encountered because there is little shared perspective. These contextual distance measures have been used in a wide variety of research.

It is this euclidean distance measure between pairs of C-score profiles which we propose to measure social distance between statuses. It is based on perspectives that incumbents of a status express. This measure facilitates studies of the relative similarity of social distance across similar statuses, changes through time, as well as studies of the consequences of differing social distances for the same positions in different organizations.

5 Social Distance In Nursing Homes: An Illustration.

Conversational interviews were conducted with 98 administrators, staff and residents from 15 selected residential settings for older people in two studies: one from a metropolitan region (McTavish and Felt, 1987) and a 1990 replication in a rural county (McTavish and Schrader, 1992). The two studies include very large and small nursing homes and retirement apartments as well as in-home care in a rural county and a variety of ownership arrangements. Given the small sample size, the findings suggested below should be viewed as suggestive rather than definitive. However,

TEXT

MINISTER: "The church is God's house where the children of God come together to praise and worship Him."

ARCHITECT: "We would like to build that church. Its maintenance costs and energy loss is low. Spaces have multiple uses. It's cost effective and impressive."

PARISHIONER: "My church is a friendly place--but it's sometimes hard to hear the sermons."

RESEARCHER: "The church is a good example of a social institution being a source of cultural mores and normative behavior."

	EMPHASIS			
Idea Category	Minister	Architect	Parishioner	Researcher
Have	-5.37	17.00	-5.37	-5.37
Being	3.18	0.28	5.49	13.55
Object	-13.84	41.54	15.70	-13.84
Sense	-2.65	-2.65	32.91	-2.65
Expression-Arena	70.78	-4.21	-4.21	-4.21
Activity	-4.01	32.14	-4.01	41.65
Science	-1.77	-1.77	-1.77	7.85
Structure	7.88	5.39	9.87	16.79
Processing Things	-6.05	9.22	18.39	-6.05

	CONTEXT			
Role	Traditional	Practical	Emotional	Analytic
Minister	25.00	-13.17	-2.49	-9.33
Architect	-8.77	25.00	-7.25	-8.98
Parishioner	4.65	-20.72	20.35	-4.28
Researcher	3.84	-1.45	-23.55	21.16

Table 1. An Illustration of E-Scores and C-Scores

systematic patterns are found, generally in hypothesized directions.

Subjects were asked, "What is it like for you around here?" and their responses (some 264,000 words) were transcribed verbatim onto a computer text file so that overall similarities and differences in perspective from each of the three positions could be examined.

5.1 Social Structure

Table 2 shows the profile of context scores for administrator, staff and resident groups. Overall, the people we interviewed describe their nursing home in emotional terms (the highest C-score in most cases), that is, in terms of personal involvement, reactions, preferences and concerns. This is much more the case for residents who are involved continuously, than for either staff or administrators for whom it is a job. As expected, administrators talk in more pragmatic terms about goal-accomplishment and achievement than is true of the other two groups. Administrators also have a more traditional perspective than the other two groups.

It is interesting to note the increased traditional and pragmatic scores, and decreased emotional

		Context Dimension			
		Traditional	Practical	Emotional	Analytic
Managers					
	Urban	2.76	11.31	5.13	-19.20
	Rural	3.80	- 3.10	19.58	-20.28
Staff					
	Urban	1.79	- .23	18.64	-20.20
	Rural	3.52	- 7.75	21.48	-17.24
Residents					
	Urban	-.03	-10.63	24.31	-13.64
	Rural	.98	- 9.93	23.97	-15.01

* Source: Metropolitan data include 11 managers, 11 staff, and 28 residents interviewed in 1984. Rural data include 4 managers, 22 staff, and 20 residents interviewed in 1990. Both studies were conducted in Minnesota.

Table 2. Social Context Score Profiles for Administrators, Staff and Residents of Urban and Rural Nursing Homes*

and analytic scores as one moves from residents to staff to administrators. This ordering fits with expectations about the intermediate position of staff in the nursing home context. The difference in context scores between rural and urban nursing homes indicates that rural managers, staff and residents are more traditional and managers and staff are less practical and more involved personally (emotional C-score), for example.

Table 3 presents the social distances of managers and staff *from* residents in both urban and rural samples. While social distances are somewhat smaller for the rural sample, the striking contrast is not the *difference* in social distance between nursing homes but the *uniformity* of social distance within a given organization. We interpret this as a result of the social structure of a given home. For example, in the urban sample, social distances between residents and managers in different nursing homes averaged 17.7 (range 50.2 points), but within the same home was only 2.1 (range 6.0 points). While different homes are quite different in the social distances, within a given home there is great uniformity, suggesting an organizational constraint: a typical distance within any given home. Different organizations seem to have different typical social distances, a phenomenon that is not adequately dealt with in organizational literature.

Finally, Figure 1 shows the pattern of social distance between these three statuses. The distances suggest a "triangular" structure: the distance between managers and residents is not simply the sum of distance between manager and staff plus staff to resident distances. For only one of our homes, the most bureaucratically structured, was the more "linear" pattern evident. Interestingly, proprietary homes have the smallest social distances between all statuses, perhaps because these structures underlie some of the typical strains felt by staff, residents and managers.

5.2 Correlates of distance between roles

Table 4 shows some distances between a) managers and staff, b) managers and residents, and c) staff and residents, within urban nursing homes for which we have complete data. Overall, managers are more distinct from residents (distance = 30.1) than they are from staff (distance =

25.1), and than staff are from residents (distance = 15.9), as expected.

Social distance measured in this way is systematically related to several organizational features, as shown in table 4. As to type of ownership, large public institutions show the largest manager to staff and staff to resident distances, while non-profit foundation-owned homes have the largest manager to resident distance. To some extent, this difference is reflected in the structure of the ownership. As expected, homes which are a part of a larger corporate structure with centralized decision-making, have a considerably larger social distance internally between the manager and both staff and residents than do homes which are not part of a corporate chain.

Correlations with the number of beds in a home show that managers in larger homes are more remote structurally and in the perspectives they share with those in other statuses. This would create a sense of remoteness and dissimilarity among statuses, and impede open discussion (what is clear expression for one becomes odd and encoded response from other perspectives). The opposite pattern is evident for staff-resident distances; the larger the home the smaller the distance. This is perhaps due to hiring practices if larger metropolitan Minnesota homes seek staff who are more similar to residents.

Nursing home reputation, as judged by outside professionals, is positively correlated with social distance and strongest for the social distance of managers from both residents and from staff. Perhaps the greater distance for higher ranked homes represents a distinctive managerial perspective or special professional training or the function of the manager in representing the home to the

		Urban Sample		Rural Sample
Manager to Resident Social Distance				
Overall Distance	mean	30.1		22.7
Between Homes	mean	17.7		13.7
	range	(50.2)		(31.7)
Within Homes	mean	2.1	1.8	
	range	(6.0)	(5.9)	
Staff to Resident Social Distance				
Overall Distance	mean	15.9		7.4
Between Homes	mean	10.2		4.2
	range	(29.3)		(14.6)
Within Homes	mean	3.6		2.4
	range	(7.2)		(8.0)

* Source: Metropolitan data on 11 nursing homes include 11 managers, 11 staff, and 28 residents interviewed in 1984. Rural data include 4 managers, 22 staff, and 20 residents interviewed in 1990. Both studies were conducted in Minnesota.

Table 3. Average Social Distance of Residents From Each Other And From Staff and Mangers Within And Between Homes*

community. Finally, manager's distance from staff is positively related to their distance from residents, with greater distance of managers from staff and residents negatively correlated with

		Manager to Staff Distance	Manager to Resident Distance	Staff to Resident Distance
Overall Social Distance		25.1	30.1	15.9
Ownership				
Proprietary (N=2)	14.2	19.9	9.9	
Church-Sponsored (N=4)	21.6	29.7	10.9	
Foundation (N=3)	24.6	42.0	20.9	
Public (N=4)		31.4	23.0	24.3
Decision-making Autonomy				
Local Decisions (N=2)		4.7	11.0	10.0
Centralized Decisions (N=8)		32.5	35.2	15.6
Correlations With Social Distance				
Number of Beds		0.28	-0.02	-0.33
Nursing Home Reputation (rank r)		0.32	0.41	0.17
Manager to Staff Distance		--	0.86	-0.28
Manager to Resident Distance		--	--	-0.20

* Source: Metropolitan Minnesota data on 11 homes, 1984.

Table 4. Relationship Between Social Distance and Selected Characteristics, Urban Minnesota Sample *

closer distance between staff and residents.

Social distance also appears to be systematically related to meanings that a nursing home has for those who work and live in it, suggesting structural effects that need further investigation. Table 5 presents nine idea (E-score) categories that illustrate some correlates of social distance in urban nursing homes. Themes like "good", "happy" and "depressed" receive more emphasis by residents than managers (E-scores for each of the nine conceptual areas are given for each status in column one of Table 5). Emphasis on the idea of "good" in staff interviews becomes more pronounced where managers are at a greater social distance from staff and *residents* (correlations of +.64 and +.63), but greater staff-resident social distance is negatively related to emphasis on "good" by both staff and managers (but essentially uncorrelated with this theme for residents). The greater the manager-staff distance, the less residents express themes of "happy", "we", and "fellow-feelings" and the more they express "depressed" themes. Also, larger distance between staff and residents is related to greater resident emphasis on "depression". Greater managerial distance seems to be negatively related to emphasis on "cognition", that is, talk of knowing or thinking, suggesting an "out of sight, out of mind" aspect of social distance. It is interesting to note the positive relationship of staff distance from residents on the cognitive concerns of both managers and residents.

The greater managerial distance from staff or residents, the more managers emphasize "community". On the other hand, the greater staff-resident distances the less emphasis on "community" by staff, suggesting that distance from residents has a different effect on manager

Idea Category	Mean E-Score	Manager to Staff Distance	Manager to Resident Distance	Staff to Resident Distance
			Correlations	
Good				
Manager	9.5	0.13	0.36	-0.22
Staff	14.4	0.64	0.63	-0.29
Resident	16.6	0.08	-0.13	0.09
Happy				
Manager	-0.1	0.12	0.26	-0.55
Staff	0.8	0.26	0.19	-0.27
Resident	3.5	-0.50	-0.35	0.10
Depressed				
Manager	-1.1	0.01	-0.20	0.16
Staff	-0.6	-0.24	-0.28	0.02
Resident	0.2	0.30	0.16	0.29
Community				
Manager	3.8	0.44	0.73	0.05
Staff	3.2	0.09	0.06	-0.22
Resident	1.6	-0.17	0.33	0.08
We				
Manager	1.8	0.07	0.15	-0.35
Staff	3.0	0.24	-0.04	0.25
Resident	1.3	-0.33	0.28	0.28
They				
Manager	4.6	0.20	0.25	0.00
Staff	6.8	-0.40	-0.38	-0.14
Resident	7.3	-0.09	-0.33	0.08
Fellow-Feeling				
Manager	3.6	-0.04	-0.28	0.30
Staff	3.0	-0.03	-0.04	0.38
Resident	1.6	-0.30	-0.27	-0.08
Implication				
Manager	4.0	0.37	0.47	0.14
Staff	6.2	-0.43	-0.25	0.42
Resident	6.0	0.67	0.74	-0.13
Cognition				
Manager	10.1	-0.62	-0.36	0.63
Staff	14.0	0.02	0.00	-0.32
Resident	10.4	-0.13	-0.23	0.34

* Source: Metropolitan Minnesota data on 11 homes, 1984.

Table 5. Correlates of Contextual Distance Between Roles And Selected Idea Category Emphasis Scores*

and staff. Greater staff distance is associated with greater resident emphasis on "we", and, likewise, the greater manager-staff distance, the more likely staff are to emphasize "we" and more likely they are to refer to "they". These indicate trends toward separation and recognition of their separate status in the organization. Managerial social distances (from staff and/or residents) seems to correlate more strongly with meanings expressed by residents than do staff-resident distances, suggesting a greater impact of the manager on meanings residents express about their setting. These relationships point up the need for further investigation of implications of social distance for meanings others express about their organization.

In sum, the manager (or staff) to resident social distance appears to be a characacteristic which distinguishes these orgranizations but is relatively uniform within a given home. The correlations lead us to expect consequences of the distances for communication and differentiation within nursing homes, for staff training and effectiveness, and for the ability of staff and managers to respond to changes in their client population. There are also implications for factors which may affect organizational "climate".

5.3 Evidence of Organizational Climate

Figure 2 presents the result of a discriminant function analysis in which selected E-scores from the conversation of managers, staff and residents are used to predict which organization they come from. The structure of the discriminant function analysis reflects differences between organizations. Three dimensions were identified (the two main ones are shown in Figure 2).

The first dimension (top to bottom) emphasizes choices or reasoning about the setting. At the top, reasoning about organizational limitations is emphasized ("Everybody had a place to sit and they didn't welcome anyone else"), and toward the bottom personal reasoning related to self and the outside ("Because we exercised together"). The second dimension from left to right suggests the basis of integration into the setting. At the left are themes emphasizing a normative integration, with a sense of guest-like attachment ("They do keep it nice and clean" and "I like the friends here"). At the right contractual integration appears to be emphasized with talk about changing arrangements and what "they" want ("So I've been able to act quickly and responsibly"--to arrange for the home). Finally, a third dimension extends toward the reader from the graph, running from an emphasis on supervised care and assistance toward relative freedom of choice at the other end of the continuum.

The fact that meaning scores permit an accurate identification of the speaker's organization suggests the existence of an organizational "climate" which pervades the speech patterns of those in each residential organization.

6 Discussion

Using computer content analysis tools, open-ended conversations with occupants of structural positions can be coded in terms of their contextual perspective. Distances between these perspectives are proposed as a measure of social distance. Such measures build upon the meanings expressed about a social context in ways which permit comparisons in a general framework of institutional themes, across social structures and through time. Thus, one could empirically raise questions about the effect of structural versus personal characteristics in altering the social distance between positions in an organization. Social distance appears to be related to structural features of nursing homes and to meanings those in different positions in the home have about their organizational situation. We suggest that contextual content analysis and other language-based measures of social distance can be explored for use in organizational research.

7 Future Prospects

The E-scores used as the basis for the foregoing analysis are similar to the use of semantic codes from a machine-readable dictionary for filtering documents for their broad subject appropriateness to a topic of interest (Liddy et al., 1993). In that study, semantic codes (such as "business," "occupations," and "political science") categorizing the use of many (but only specialized) words into 124 major fields are used. After processing a text to disambiguate words (including stemming, part of speech tagging, using sentence-level context-heuristics, analyzing co-occurrence probabilities with a correlation matrix), a vector of frequencies is generated and normalized (to control for document length). A query is coded using the same technique and is matched with semantic vectors characterizing a document collection based on a computed predicted similarity value. (The document collection has also been clustered using agglomerative clustering algorithms so that cluster centroids can be compared to the query to provide a searcher with a "semantically cohesive cluster" for browsing.) This approach has been found to provide enhanced information retrieval results.

As noted above, E-scores provide a mechanism for identifying a speaker's organization. To understand why this mechanism is similar to the use of semantic codes, we have begun a semantic analysis of the words in MCCA's idea categories (as part of a longer-term effort to extend the words and categories that can be used and to port the system to a different computing environment). MCCA's dictionary consists of 11,000 distinct words, with an average of about 95 words per category. Using DIMAP-2 (CL Research, 1992) and DIMAP-3 (CL Research, 1995), the MCCA dictionary was uploaded into machine-tractable form for study with other lexical resources. In particular, sublexicon dictionaries were created for individual idea categories by converting WordNet™ 1.4 (Miller et al., 1993) entries for the words in the category into DIMAP format. The hypernymic, hyponymic, and other relations in WordNet could then be examined to assess the semantic characteristics of these sublexicons.

WordNet is a semantic network of about 100K words grouped into "synonym sets" (synsets) of about 5 words or phrases. These synsets are connected by various semantic relations with one another, identifying more general and more specific concepts, part-of relations, synonymic relations, antonymic relations, etc. These are among the more general relations that exist among words as identified in lexical semantics (Cruse, 1986 and Nida, 1975). Moreover, the hierarchical structure of WordNet is such that all nouns and verbs are grouped into about 150 semantically coherent categories. In examining the MCCA sublexicons, after removing some senses from the WordNet entries that were clearly not members of a particular idea category, an (intuitively) high degree of coherence was found. In particular, most of the words were connected to other words in the sublexicon by morphological and derivational relations, taxonomic relations, and partitive relations, thus containing common semantic components and making them semantic domains. In addition, using the WordNet data, it was possible to extend the words that might be associated with each idea category, primarily by identifying hyponyms (narrower terms).

These relations provide a stronger unity than do the subject categories of a dictionary, which only indicate specialized usages of particular words. These subject categories do not capture semantic components which reflect fine-grained meanings inherent in particular words. It may be suggested that representing texts with such meanings will provide better characterizations of texts and result in even better information retrieval. Krovetz and Croft (1992) have already shown that using such derivational and other lexical relations results in improved identification of word senses in disambiguation with an accompanying improved retrieval, as measured by improved recall and precision.

The use of WordNet and other lexical resources allows the further refinement of the idea categories into semantic components. Considerable research is now identifying such components (as well as features characterizing stylistic register, other pragmatic information, and words that are associated with particular functions--such as expressive, interpersonal, and emotive--of language) and incorporating them into machine-readable dictionaries. In fact, it is possible that the functions of language may correspond to the contexts set up in MCCA. The effect of this work on the content analysis of the present study of nursing homes will be examined next; it is expected that this will lead to sharper and better-defined characterizations of the results reported herein. These results are likely to support other similar analyses using MCCA contextual content analysis on numerous other kinds of texts.

8 References

Cleveland, C.E., McTavish, D. & Pirro, E.B. (1974): Contextual Content Analysis. In: Proceedings of the ISSC/CISS Workshop on Content Analysis in the Social Sciences, a conference sponsored by the Standing Committee on Social Science Data of the International Social Science Council, UNESCO, Centro Nazionale Universitario del Calcolo Eletronico (CUNCE), Pisa, Italy September 5-13.

CL Research (1992): DIMAP-2 (Dictionary Maintenance Programs) User's Manual. Gaithersburg, MD: CL Research.

CL Research (1995): DIMAP-3 (Dictionary Maintenance Programs) User's Manual. Gaithersburg, MD: CL Research. (In preparation.)

Cruse, D. A. (1986): Lexical Semantics. Cambridge: Cambridge University Press.

Kadushin, C. (1962): Social Distance Between Client and Professional. American Journal of Sociology 67(5): 517-531.

Kidwell, I. J. & Booth, A. (1977): Social Distance and Intergenerational Relations. Gerontologist 17(5): 412-420.

Krovetz, R. & Croft, W. B. (1992): Lexical Ambiguity and Information Retrieval. ACM Transactions on Information Systems 10(2): 115-41.

Liddy, E. D., Paik, W., & Yu, E. S. (1993): Document Filtering Using Semantic Information from a Machine Readable Dictionary. In: Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives, sponsored by the Association of Computational Linguistics, Ohio State University, Columbus, Ohio, USA, June 22, 1993: 20-29.

McPherson, J. M. & Smith-Lovin, L. (1987): Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. American Sociological Review 52: 370-379.

McTavish, D. G. & Felt, D. (1987): Role Distances and Atmosphere in Nursing Homes. Sociology of Rural Life 9(1): 3-4,8.

McTavish, D. G. & Pirro, E. B. (1990): Contextual Content Analysis. Quality and Quantity 24: 245-265.

McTavish, D. G. & Schrader, S. L. (1992): Residential Options for Rural Minnesota Elderly. Sociology of Rural Life 12: 3-4,7-8.

Miller, D. C. (1983): Handbook of Research Design and Social Measurement, 4th edition, New York: Longman 329-333.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. & Tengi, R. (1993): Five Papers on WordNet™. CSL Report 43. Princeton, NJ: Princeton University Cognitive Science Laboratory.

Nida, E. A. (1975): Componential Analysis of Meaning. The Hague: Mouton.

Reiss, A. J., Jr. (1961): Occupations and Social Status, New York: Free Press of Glencoe.

9 Biographical Sketches

Donald McTavish, PhD, is a Professor of Sociology at the University of Minnesota. Areas of interest include computer content analysis, research methods and statistics, and the sociology of age.

Kenneth Litkowski is the owner of CL Research, which markets software for use in developing and maintaining lexicons for natural language processing. He uses this software for research in the semantic structure of language.

Susan Schrader, PhD, is Assistant Professor of Sociology at Augustana College. Areas of interest include institutional settings for older people and the sociology of age.