



Chen, Guocai and Warren, Jim and McArthur, Robert and Bruza, Peter D. and Kralik, Debbie and Price, Kay (2007) Understanding Individual Experiences of Chronic Illness with Semantic Space Models of Electronic Discussions. In *Proceedings Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007. CBMS '07*, pages pp. 548-553, Maribor, Slovenia.

© Copyright 2007 IEEE

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Understanding Individual Experiences of Chronic Illness with Semantic Space Models of Electronic Discussions

Guocai Chen¹, Jim Warren¹, Robert McArthur², Peter Bruza³, Debbie Kralik⁴ and Kay Price⁵

¹ Department of Computer Science – Tamaki, University of Auckland, New Zealand

² CSIRO, Canberra, Australia

³ Faculty of IT, Queensland University of Technology, Brisbane, Australia

⁴ Royal District Nursing Service, Adelaide, Australia

⁵ School of Nursing & Midwifery, University of South Australia, Adelaide, Australia
g.chen@auckland.ac.nz

Abstract

Electronic discussion groups provide a convenient forum for individuals to share their experiences of chronic illness. The language use of individual participants, and the way their language shifts over time, may provide implicit indications of important shifts in sense-of-self. This paper relates experience with application of the Hyperspace Analogue to Language (HAL) model for automatic construction of a dimensional model from a corpus of text. HAL is applied to 17 months of discussion on a closed list of 20 women coping with chronic illness. The discussion group was moderated for a focus the phenomenon of 'Transition' - how people can learn to incorporate the consequences of illness into their lives. The current phase of research focuses on identification of clusters of words that can represent key aspects of Transition. The HAL models for two participants have been analyzed by experts in Transition to form candidate clusters. These clusters are then used as a basis for contrasting the language usage of an individual participant over time as compared to the entire corpus. We have not yet found a reliable basis for identifying transitions in an individual based on their entries into a discussion forum, although the clusters may have some inherent value for introspection on individual experiences and Transition in general. We report challenges for interpretation of the HAL model related to the correlation of dimensions and the impact of group dynamics.

1. Introduction

Internet communities provide a forum for people to seek and share information, validate and affirm the illness experience and connect with others [1]. Participants of electronic discussion groups provide emotional support, guidance and health information; support self education and self responsibility; encourage patients' initiative and assertiveness; and provide members with an opportunity to help others [2]. Such groups constitute a distinct and complementary way for health consumers to empower themselves via Information Technology as compared to less social computer use, such as interrogating search engines to find online health information [3]. Moreover, the Internet provides a useful venue for interaction for people whose illness may make it difficult for them to manage regular face-to-face meetings [4].

Chronic illness can impact significantly on people's lives. Electronic discussion groups provide a venue for individuals to discuss the 'transitions' in their lives as a result of chronic illness. By Transition we mean the ways people can learn to incorporate the consequences of illness into their lives [5]. A working definition of Transition is: a process of convoluted passage during which people redefine their sense of self and redevelop their self agency in response to life events such as chronic illness. We have been examining the benefits of

membership of electronic discussion groups, and the data generated, to better understand Transition and the way sense of self is reflected in the language people use when communicating with others in the discussion group [4, 6]. Transition can be considered as involving several distinct stages that are iteratively revisited as illness imposes new barriers to the undertaking of ordinary life (see figure 1).

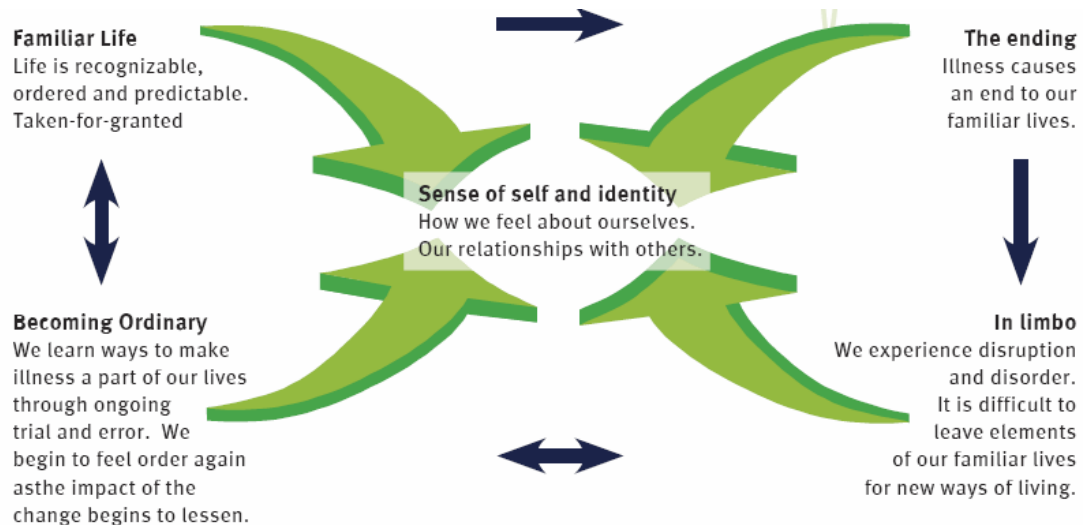


Figure 1. Stages of Transition in chronic illness (from [7])

Hyperspace Analogue to Language (HAL) is a model for automatic construction of a dimensional model from a corpus of text [8]. In the case of HAL, an $N \times N$ matrix is instantiated with an N -length vector for each unique word occurring in a corpus. A 'window' several words in width is moved across the corpus; wherever two words occur within the window the value at their intersection in the matrix is incremented. Thus, a corpus is converted to a high-dimensional semantic space, with minimal consideration to grammar. Other significant semantic space models include LSA [9], which is widely used for document indexing, and the more recent COALS model [10]. We were particularly influenced to use HAL because it has some track record in representing emotion in text. Burgess and Lund [11] examined whether HAL could represent abstract concepts, such as love, hate, joy. They found that, in a comparison with human raters in predicting abstract variables for a set of words, "global co-occurrence information carried in the word vectors can be used to predict a tangible proportion of the human likert scale ratings."

Herein we explore extended application of HAL in an attempt to detect the language of Transition used by the participants of an electronic discussion forum.

2. Method

Our corpus for analysis is based on the long-term conversations between 20 women who self-describe as struggling to cope with chronic illness. These women participated in an electronic discussion group implemented as a limited-access majordomo email list service, facilitated by an expert in Transition (DK) who moderates the conversation and can contact the participants outside of the list if needs arise. We have previously analyzed data from this list [6] to examine the quantitative changes in language use on dimensions of 'kin' and 'negative emotion' (which were expected to roughly correspond with the key Transition concepts of

‘ordinariness’ and ‘extraordinariness’). The HAL vectors for ‘sense of self’ (the sum of HAL vectors for words ‘me’, ‘my’, ‘I’, and ‘myself’) for two participants are used in the present study. Seventeen months of data from the corpus are used for further analysis presented herein.

The present study has two major phases. First, a Semantic Spaces researcher (JW) facilitated two Transition experts (DK, KP) to cluster the 200 largest-magnitude words in the sense-of-self vector for each of two participants in the discussion group by sequentially placing the words on a large surface. The experts were encouraged to create labels for the clusters whenever they saw a natural association emerge. These cluster maps were then used to identify word clusters to form axes for review of participant data over time. The expectation is that major Transitions (a la the arrows in figure 1) should show as measurable changes in the relationship of the participant data to axes defined by key concept word clusters. For this purpose the strength of relationship, r , of a participant, P , in a given month to a given concept cluster, C , is defined as the projection of the participant’s text for that month onto the text of the cluster:

$$r_{\text{month}}(P,C) = \text{cosine}(W_{P,\text{month}}, W_C) \quad (1)$$

where $W_{P,\text{month}}$ is a vector that is the sum of the HAL vectors for all the words a participant used in a given month based on the corpus of that participant’s emails for that month, and W_C is the sum of the HAL vectors of the concept cluster words across the entire corpus; both $W_{P,\text{month}}$ and W_C are of length N , where N is the total vocabulary size in the corpus. If a participant did not use a given word in a given month, the corresponding element of $W_{P,\text{month}}$ has a value of zero.

3. Results

The cluster maps for two participants (‘Susan’ and ‘Cyndi’, not their actual names) were formulated by expert clustering of their sense-of-self HAL vectors (figures 2 and 3, respectively). Promising segments of Susan’s map (marked on figure 2) have been used to define axes for projection for further analysis of Susan’s transitions in language usage over time as she participated in the forum. Table 1 shows the 20 largest HAL values and their associated words for these three new axes as well as the two axes used in [6]. Figure 4 shows the projection, as per equation 1, of Susan’s submissions to the discussion forum by month onto these five axes. Susan’s stages of Transition were assessed by a Transition expert who moderated the forum (DK) based on qualitative analysis of her text entries to the forum. The projections in figure 4 map poorly to those identified states.

4. Discussion

Manual clusters of ‘sense of self’ HAL vector terms provide interesting pictures of an individual’s participation in a discussion forum. The Transition experts clustering the terms found them representative both of Transition and of the character of the individuals on which they were based. Nuances of the discussion appear in the HAL vectors (note the word ‘Spoons’ in figure 3, which sources from a metaphor discussed in the forum comparing limited energy for daily tasks to a budget of a handful of spoons). There are notable contrasts between the cluster maps for different individuals (figure 2 v. figure 3). The participant Susan was able to meet face to face with the researchers and view her cluster map. She felt that it was highly representative of her thoughts during that period of her life. Subsequent use of the cluster map to plot Susan’s transitions, however, has not produced a viable correspondence with expert assessment of the key transition points. In fact, all the axes are (as visible in figure 4) highly correlated with one another, and in fact are positively correlated with the amount of message traffic Susan generated during a given month.

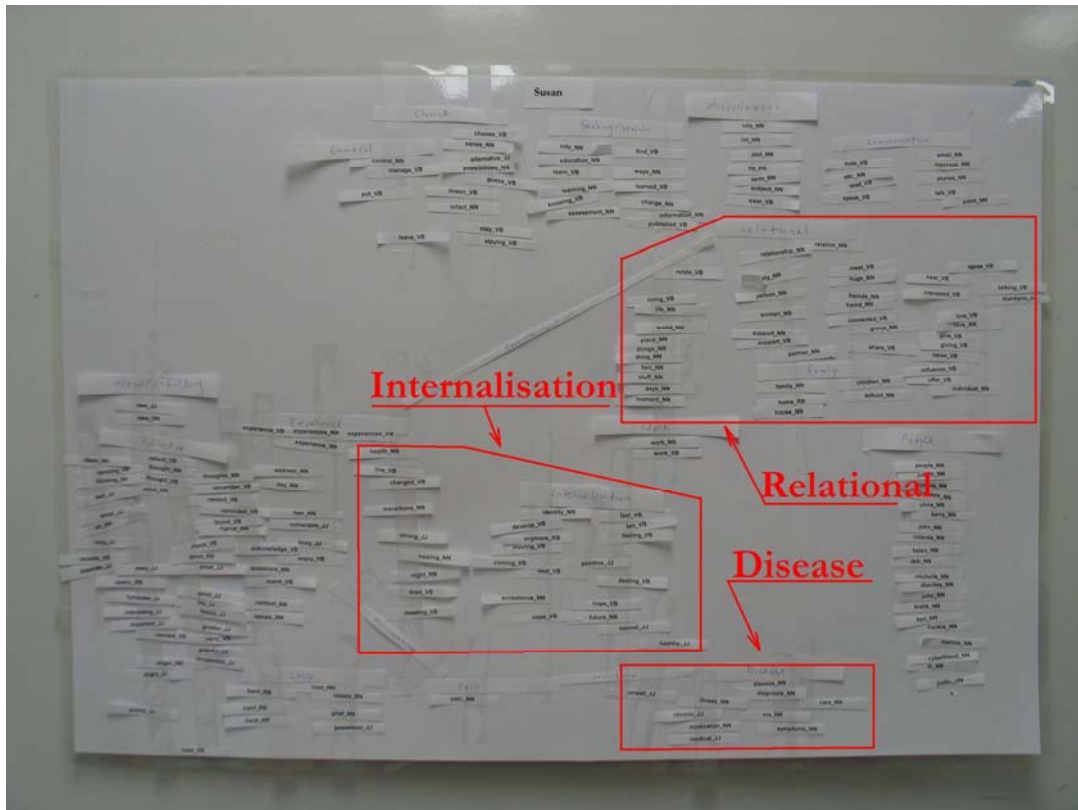


Figure 2. Manually-produced cluster map of largest sense-of-self terms for Susan (overlaid with subsequently selected projection axes)



Figure 3. Segment of manually-produced cluster map for Cyndi

Table 1. Largest values for HAL vectors of projection axes

<u>Kin</u>	<u>Negative Emotion</u>	<u>Internalisation</u>	<u>Disease</u>	<u>Relational</u>
time 53.6	pain 711.9	time 247.81	illness 808.3	good 192.86
friends 40.4	fatigue 306.5	feel 238.88	chronic 624.6	time 190.81
good 35.8	back 255.75	good 227.12	people 341.4	people 187.44
children 34.4	chronic 230.0	people 202.31	living 323.3	michelle 166.61
mother 32.1	time 188.0	hope 198.27	life 262.7	illness 148.47
work 31.3	feel 183.75	things 193.12	pain 261.4	things 144.61
years 30.9	day 183.0	illness 192.38	experience 189.9	life 137.72
life 30.4	hope 179.5	life 181.65	feel 186.7	feel 134.56
told 28.1	life 167.88	pain 176.65	time 166.5	hayden 129.5
family 27.9	bad 165.88	day 149.77	health 163.5	life's 126.97
back 27.8	experience 147.9	feeling 147.15	good 163.4	adam 118.5
things 27.0	things 144.88	back 146.69	person 155.0	hope 117.5
day 26.3	people 144.0	lot 130.65	things 141.9	precious 113.06
home 26.3	feeling 138.75	di 120.15	long 125.9	day 111.36
people 25.7	lot 137.88	michelle 119.0	important 118.9	mum 110.33
feel 24.2	worse 132.88	bit 117.69	back 111.8	back 108.86
hope 23.1	good 127.75	today 114.54	find 109.0	hanna 108.39
wife 22.6	find 121.38	chronic 113.54	condition 105.9	chronic 107.33
daughter 21.8	sleep 111.63	find 107.12	changed 103.5	work 105.02
working 21.5	days 105.5	make 105.85	lot 101.8	pain 100.5

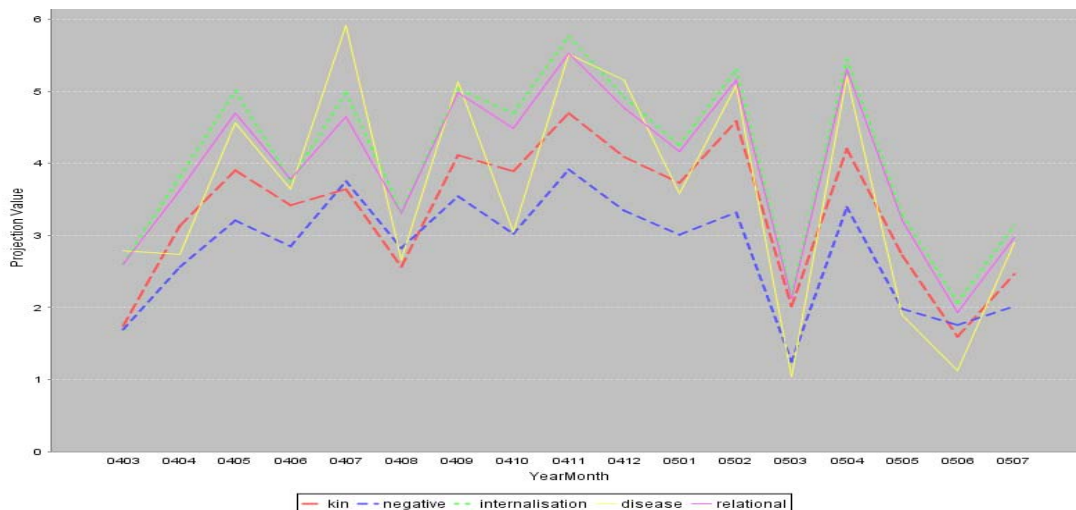


Figure 4. Projection of Susan's discussion entries by month

There have been some promising results with semantic spaces, and HAL in particular, in recent years [11, 12, 13]. These studies have shown that HAL can be used to extract knowledge directly from email utterances, and that singular value decomposition (SVD; the basis of LSA) can be applied with HAL to extract tacit knowledge. The studies have also shown the validity of using a model like HAL, originally developed on a 300 million word corpus from Usenet, on datasets of a few sentences [13] to a few thousand emails or documents [12]. The analysis as reflected in figure 4 has made some simplification, notably the omission of part-of-speech tagging (evident in the earlier phase by markers such as "NN" for noun in figures 2 and 3) and no application of SVD. It is not thought that these simplifications account for the patterns in figure 4, but the potential benefit of SVD in particular warrants further explanation.

Some of the challenge is accountable by the nature of the corpus. Participants did not always

simply state how they were feeling at the present time. Consider the following snippet where Susan is responding to a question put to the forum by the moderator:

>> 1. Can you talk about the relevance of your attitudes and values to the
>> experience of living with chronic illness?

When I came "down" with MS I wasn't really surprised I'm not really sure where that came from, There was a deep familiarity with what I was experiencing, a knowing "of" the experience. Had I dreamt it, was it from a previous life, where there signs that had been appearing for sometime that I hadn't connected together, was there an emotional process that I was experiencing that in retrospect mirrors my now current physical experience. Does my interpretation of the disease, A hardening process of the nerves relate to my lived experience of loss and grief from an early age???

Our filtering ignores the quoted text (marked by ">>" above) but cannot ignore the fact that the participant, viewed as having transitioned to Familiar Life through Becoming Ordinary (see figure 1) by the time of this entry, has taken the occasion to reminisce and hence is using language more representative of her earlier states.

5. Conclusions

Online forums are a valuable resource for people to share tacit knowledge about how to live with chronic illness, to provide mutual support, and express and explore one's feelings. We have found the HAL dimensional semantic space model to provide thought-provoking representations of the language used in such forums, but have not yet been able to quantitatively identify forum participants' transitions in sense-of-self. Progress in this area will be valuable both to support further understanding of the experience of chronic illness but also has potential to provide monitoring tools to support forum moderators.

6. References

- [1] Eysenbach, G., et al., Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *British Medical Journal*, 2004. 328(7449): p. 1166.
- [2] Ferguson, T., From patients to end users. *British Medical Journal*, 2002. 324(7337): p. 555-556.
- [3] Warren, J., IT-Empowered consumers and advocates for chronic disease management. *Health Care and Informatics Review Online*, September 2006 (URL: <http://hcro.enigma.co.nz/website/index.cfm?fuseaction=articledisplay&FeatureID=130906>)
- [4] Kralik D, Price K, Warren J, Koch T. Issues in data generation using email group conversations for nursing research. *J Adv Nurs*. 2006 Jan;53(2):213-20.
- [5] Kralik, D., The quest for ordinariness: transition experienced by midlife women living with chronic illness. *Journal of Advanced Nursing*, 2002. 39(2): p. 146-154.
- [6] McArthur, R., Bruza, P., Warren, J., Kralik, D., Projecting computational sense of self: A study of transition in a chronic illness online community. In: *Proceedings of the 39th Hawaii International Conference on Systems Sciences*, IEEE Computer Society Press, January 2006.
- [7] Kralik, D., van Loon, A., Telford, K., *Transitions in Chronic Illness: Booklet 11: Understanding Transition*. University of South Australia / Royal District Nursing Service. 2005. URL: http://www.rdns.org.au/research_unit
- [8] Burgess, C., Livesay, K., Lund, K., Explorations in context space: words, sentences, discourse. *Discourse Processes*, 1998. 25: 211-257.
- [9] Landauer, T., Foltz, P.W., Latham, D., Introduction to latent semantic analysis. *Discourse Processes*, 1998. 25: 259-284.
- [10] Rohde, D., Gonnerman, L., Plaut, D., An improved method for deriving word meaning from lexical co-occurrence, 2004. URL: <http://dlt4.mit.edu/~dr/COALS/Coals.pdf>
- [11] Burgess, C., Lund, K., Representing abstract words and emotional connotation in a high-dimensional memory space. *Cognitive Science Proceedings*, 1997. LEA. pp. 61-66. URL: [http://hal.ucr.edu/pdfs/Burgess_Lund\(1997b\).pdf](http://hal.ucr.edu/pdfs/Burgess_Lund(1997b).pdf)
- [12] McArthur, R., Bruza, P. Discovery of implicit and explicit connections between people using email utterance. In: *8th European Conference on Computer-Supported Cooperative Work (ECSCW)*. 2003. Helsinki, Finland: Kluwer.
- [13] McArthur, R., Bruza, P., Discovery of tacit knowledge and topical ebbs and flows within the utterances of online community. In: *Chance Discovery*, Y. Ohsawa and P. McBurney, Editors. 2003, Springer, pp. 115-131.