

Using Entropy to Measure Design Creativity

Using a text Based Analysis Tool on Design Protocols

Jeff W.T. Kan and John S. Gero
Krasnow Institute for Advanced Study, USA
kan.jeff@gmail.com; john@johngero.com

This paper presents a case study examining Shannon's entropy as a tool for measuring design creativity from protocols. Two design sessions were analyzed. One of the sessions was judged to be more creative based on their outcomes and qualitative analyses. We calculated the text entropy of the two protocols with the hypothesis that the creative session will have a higher text entropy. However their relative text entropies were the same. A further analysis of the two sessions was carried out by accounting for the semantics of the protocols. The protocols were segmented and segments containing words of similar meaning were linked. These links formed a linkograph. The entropy of the linkographs of the two sessions were than measured. The creative session has a higher linkograph entropy. This result indicates the potential of using entropy as a measure of design creativity.

Keywords: Design creativity, protocol analysis, entropy, linkograph, cognitive synonym

1. Introduction

This paper investigates a text-based measurement tool to study design creativity, where the text is derived from a design protocol. The focus is on entropy – a measurement of the amount of information in the text; a low value of entropy indicates no unexpectedness or surprise. We assume creativity is related to unexpectedness or surprise, along with novelty and usefulness, which are not measured here. It commences by examining Torres' [1] claims – “one can easily conclude that the entropy is indeed a very useful measure to compare the richness of two or more texts” [1, p1] – with the assumption that design creativity is related to the richness of a design protocol. We used two design sessions' transcripts to test Torres' claim. Text entropy only looks at the number of different words without including their meaning or their associations. Linkography [2], on the other hand, studies the connection or association of ideas based on the design protocol. It has been used as a tool to compare design productivity and study creativity. The authors, in a previous study, proposed a way of using entropy to measure the linkographs [3, 4, 5]. This measurement adopts the position that a creative process embraces both integration and diversification of ideas. When there is no diversification of ideas in the process the likelihood of being creative is low. On the other hand, if all ideas are totally different with no association the likelihood of producing a final design is low.

2. Entropy and the Richness of Text

Claude Shannon, the founder of information theory, suggested a measure of information associated with a communication source, called entropy [6]. He suggested that the amount of information carried by a message is based on the probability of its occurrence. If there is only one possible outcome, then there is no need to communicate additional information because the outcome is known. To illustrate this with a simple example, consider transmitting a piece of information consisting of ten ON|OFF signals and one of them is OFF but the others are ON. The probability of an OFF symbol, $p(\text{OFF})$, is 0.1 and the probability of an ON symbol, $p(\text{ON})$, is 0.9. Consider the following two cases:

1. If the first signal the receiver gets is an OFF symbol ($p=0.1$), then no further transmission is required as the following signals carry no additional information. This, a stochastic process, assumes that the receiver knows the total number of signals (10), the probabilities of the symbols (ON/OFF), and that the total probability equals 1 ($p(\text{ON})+p(\text{OFF})=1$).

2. If the first signal being transmitted is an ON symbol ($p=0.9$), then the receiver is uncertain of the value of the next signal. Transmission is still required to complete the information.

Shannon [2] proved and derived the entropy H , the average information per symbol in a set of symbols with *a priori* probabilities, is:

$$H = - \sum_{i=1}^n p_i \log(p_i) \text{ with } H = \sum_{i=1}^n p_i = 1 \quad (1)$$

where p_i are the probabilities of occurrence of each symbols

Torres [1], using equation (1), defined the entropy of a text T with λ words and n different ones by:

$$H_T(f_1, \dots, f_n) = \frac{1}{\lambda} \sum_{i=1}^n f_i [\log_{10}(\lambda) - \log_{10}(f_i)] \quad (2)$$

where $f_i, i = 1, \dots, n$, is the frequency of each i -word in the text T .

In this equation, with the same number of words in a text, a higher variety of words will yield a higher text entropy. To compare texts with different numbers of words λ , a kind of “relative entropy” H_{rel} is introduced [1]. It is defined as the quotient between the entropy H_T of the text and the maximum entropy H_{max} , and multiplying by 100 as a percentage:

$$H_{rel} = \frac{H_T}{H_{max}} \times 100 \quad (3)$$

The maximum entropy H_{max} occurs when the text has all different words, i.e. the entropy of a text with the same number λ of words in which each word occurs only once (i.e., $n = \lambda, f_i = 1$). With equation (1) it becomes [1]:

$$H_{max} = \frac{1}{\lambda} \sum_{i=1}^n 1 [\log_{10}(\lambda) - \log_{10}(1)] = \log_{10}(\lambda) \quad (4)$$

To test if text entropy can measure the richness of text we measured the lyrics of two famous songs: *The Sound of Silence* and *Twinkle Twinkle Little Star*. We labeled them as Song A and Song B respectively in Table 1. We expect song A will be richer than the nursery song B. With equations (2), (3) and (4) we can calculate the text entropy and the relative entropy of the two songs. Torres [1] provided the source code for the calculations.

Table 1. Extract of the lyrics and their entropy measurements

Song A	Song B
Hello darkness, my old friend,	Twinkle, twinkle, little star,
I've come to talk with you again,	How I wonder what you are.
Because a vision softly creeping,	Up above the world so high,
Left its seeds while I was sleeping,	Like a diamond in the sky.
And the vision that was planted in my brain	Twinkle, twinkle, little star,
Still remains. Within the sound of silence.	How I wonder what you are!
Word count, $\lambda = 40$	Word count $\lambda = 32$
Relative Entropy, $H_{rel} = 96$	Relative Entropy, $H_{rel} = 84$
Text Entropy, $H_T = 1.5$	Text Entropy, $H_T = 1.3$

The lyric of the whole song, A has a word count of 217, text entropy of 1.9 and relative entropy of 82. Song B has a word count of 109, text entropy of 1.5 and relative entropy of 76. These measurements match our expectations. Song A has higher text entropy and relative entropy than song B. From this example, one of potentially many, we hypothesise that design creativity is related to the richness of the transcription of a design protocol.

3. The Design Sessions

In this study, data were obtained from the CRC for Construction Innovation project [acknowledgment]. In that project *in-vitro* studies were conducted with five pairs of architect designers in three different collaboration settings – face-to-face, Internet shared drawing-board and 3D virtual world. Two face-to-face sessions, labeled A and B, were used in this study out of the 15 sessions. The designers were asked to design a contemporary art gallery. They were given 30 minutes to generate a conceptual design. Judging by the outcomes, a creative session was selected together with a less creative session.

3.1. Qualitative Analysis of the Two Sessions

Session A can be divided into four stages or episodes, based on the design activities. In the first episode they dealt with the requirements and site (about 3.5 minutes); in the second episode they analyzed, planned and developed concepts in the plan (Figure 1a; about 9 minutes); in the third episode they developed the 3D form in elevation (Figures 1b, c and d; about 9 minutes); and in the final episode they worked on the layout in the plan until the end (8.5 minutes), but they did not finish it within the 30 minutes allocated for the session. They produced six sheets of drawings. This session can be qualitatively assessed a creative session.

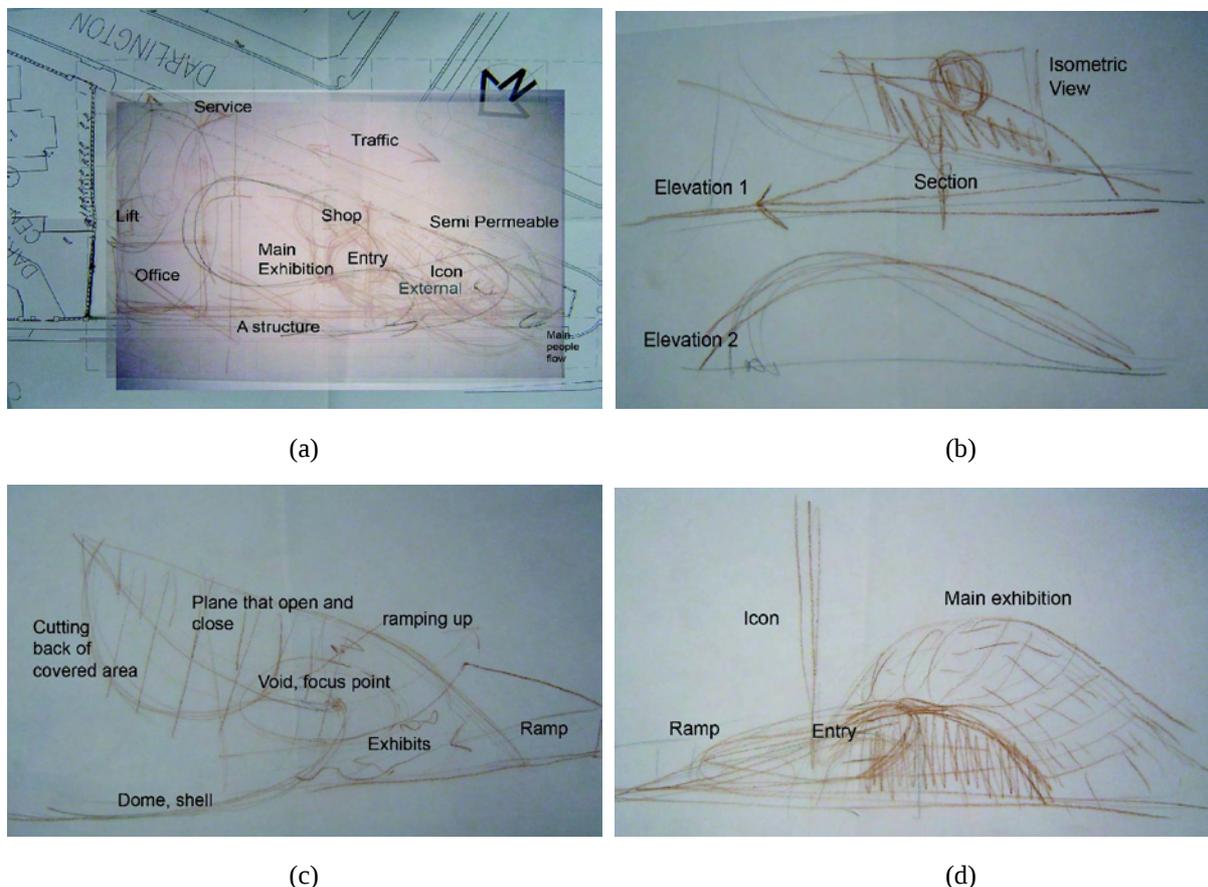


Figure 1 (a) The first two sheets of drawing. (b) Section, elevation and 3D view of the proposed gallery. (c) The

plan of an organic shaped building. (d) Elevation of the gallery.

In session B, the designers only worked on the plan. They spent about 4 minutes studying the brief without verbalizing, and then used another 5 minutes understanding the required areas in relations to the site coverage; in particular the location of external exhibition. In about ten and half minutes they discussed the location of the cafe and kitchen. At around 14 minutes into the session one designer found that the cafe was not in the requirements. Location of service dock, entrance, workshop and store were discussed, Figure 4. At around 17 minutes the location of the entrance along with its glazing was proposed. After that they worked out the location of stairs, toilets, offices together with its size in the second level, Figure 2. They did not finish the design in the time allocated. This session can be qualitatively assessed as a pragmatic session.

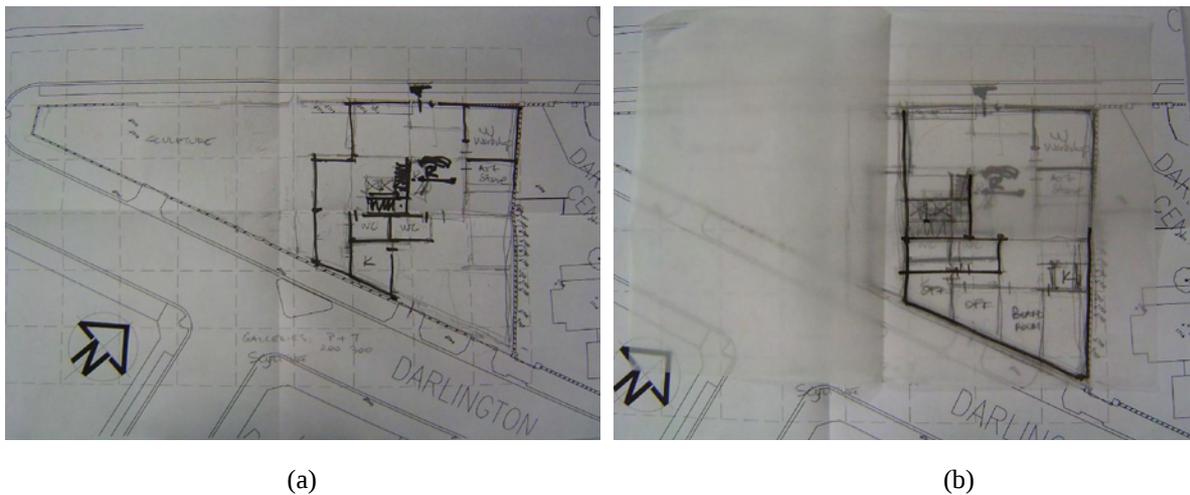


Figure 2 (a) Ground floor plan. (b) Second floor plan

4. Text Entropies of the Sessions

Table 2 is the results of measuring the text entropy of the transcripts of the two sessions. The main difference is the word count. They have the same relative entropy and the “creative” session has a slightly higher text entropy.

Table 2, The word count, relative entropy and text entropy of the two sessions

Transcript A (Creative)	Transcript B (Pragmatic)
Word count, $\lambda = 4563$	Word count $\lambda = 2309$
Relative Entropy, $H_{rel} = 64$	Relative Entropy, $H_{rel} = 64$
Text Entropy, $H_T = 2.3$	Text Entropy, $H_T = 2.2$

These results suggest that text entropy is not a good measurement tool for creativity.

5. Linkography

Linkography – a technique used in analyzing design protocols – has been reported to reveal the quality of a design process [2, 7] and the creativity of the ideas [8, 9]. A linkograph is constructed by segmenting design protocols into smaller units called a “move” and then connecting them by an analyst using domain knowledge and commonsense. Sequential “moves” are placed along the horizontal axis. When two “move” are related they are joined by a “link”. The second column of Table 3 shows some examples of linkographs. The design process can then be examined in terms of the patterns of move associations. Goldschmidt [2] identified two types of links: backlinks and forelinks. Backlinks are links of moves that connect to previous moves. Forelinks are links

of moves that connect to subsequent moves. Conceptually forelinks and backlinks are very different. “Backlinks record the path that led to a move's generation, while forelinks bear evidence to its contribution to the production of further moves” [7 p. 196]. The link index and critical moves were devised as indicators of design productivity. A link index is the ratio between the number of links and the number of moves. Critical moves are design moves that are rich in links. [2, 7]. Design productivity is positively related to the link index and critical moves; higher values of link index and critical moves indicate a more productive design process. Later, Goldschmidt [8] provided empirical evidence that quality outcome, creativity, hinges on good ideas or what she called critical moves. der Lugt [9] used the same method, with some extensions to trace the design idea generation process and empirically verified the correlation between creative qualities of ideas and the well-integratedness of those ideas.

5.1 Entropy of Linkographs

Kan and Gero [3] proposed an approach based on Shannon's information theory, to measure linkographs. They suggested that a rich idea generation process is one where: (1) the structure of ideas is reasonably integrated and articulated, and (2) there is a variety of moves. They argued that an empty linked linkograph can be considered as a non-converging process with no coherent ideas and a fully linked linkograph represents a fully integrated process with no diversification [4].

Kan and Gero [5] suggested, based on the conceptual difference between forelink and backlink, also to measure entropies in rows according to “linked” or “unlinked” as in Figure 3. Horizonlink (Figure 3 c) carries the notion of distance/time between the linked moves.

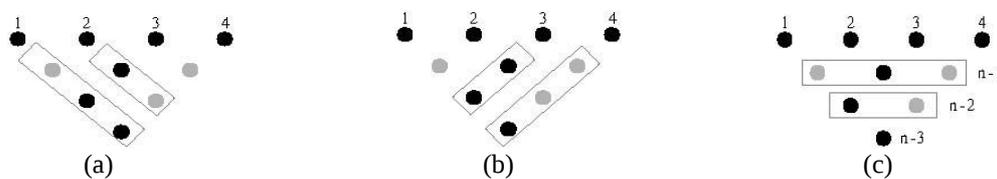


Figure 3. Abstracted linkograph for entropy measurement, back dots denote “linked” between moves and gray dots denote “unlinked”.(a) measuring entropy of forelinks of each row, (b) measuring entropy of backlinks of each row, and (c) measuring entropy of horizonlinks [4].

Using “linked” and “unlinked” as the symbols, the probability of “linked”, $p(linked)$, will be the frequency (or number) of “linked” nodes divided by the total number of nodes in that row. Similarly, the probability of “unlinked”, $p(unlinked)$, will be the number of “unlinked” nodes over the total number of nodes in that row.

There are only two symbols, putting their probabilities in equation (1) the entropy of each rows become:

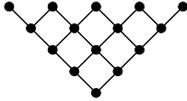
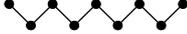
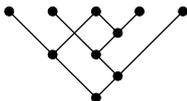
$$-p(linked)\text{Log}(p(linked)) - p(unlinked)\text{Log}(p(unlinked)) \text{ where } p(linked) + p(unlinked) = 1 \quad (5)$$

Kan and Gero [4] suggested that “forelink entropy measures the idea generation opportunities in terms of new creations or initiations. Backlink entropy measures the opportunities according to enhancements or responses. Horizonlink entropy measures the opportunities relating to cohesiveness and incubation” [4 p.335-336].

Table 3 shows four hypothetical cases with five design moves together with their cumulative entropies. The cumulative entropies are the summation of forelink, backlink and horizonlink entropies of all rows.

Table 3. Hypothetical linkographs, their interpretations and their entropies [4].

Linkographs	Interpretations	Entropy
Case 1 • • • • •	Five moves are totally unrelated; indicating that no converging ideas, hence very low opportunity for idea development.	0

Case 2		All moves are interconnected, this shows that this is a total integrated process with no diversification, hinting that a pre-mature crystallization or fixation of one idea may have occurred, therefore also very low opportunity for novel idea.	0
Case 3		Moves are related only to the last one. This indicates the process is progressing but not developing indicating some opportunities for ideal development.	5.46
Case 4		Moves are inter-related but also not totally connected indicating that there are lots of opportunities for good ideas with development.	8.57

5.2 Words, Creativity, and Forelink Critical Moves

Among the three measurements (λ , H_{rel} and H_T) from Section 3.2, the word count is the only difference between the creative and the pragmatic session. Table 4 documents the number of segments (moves) and the average number of words per segment. The natural turn-taking was used as the segmentation method.

Table 4. Word count and segment count of both sessions

A (Creative)	B (Pragmatic)
4563 words	2309 words
413 segments, 11 words per segment	252 segments, 9.2 words per segment

A word count study of the team protocol data from the 1994 Delft Protocols Workshop [10] was used to compare with a previous study [7] that indicates creativity. Table 5 summarizes the percentage of three individual's critical moves with 7 links (CM^7) and forward link critical moves with 7 or more links ($f-CM^7$). $f-CM^7$ was believed to indicate creativity [7]. Figure 4 shows the relationship between critical moves and word counts.

Table 5. Comparing critical moves and word counts of three individuals from study described in [7] (%)

	Ivan	John	Kerry
CM^7	24.3	43.3	32.4
$f-CM^7$	28.6	42.8	28.6
% of words	30.27	44.96	22.63
Words per segmentation	7.85	11.65	8.25

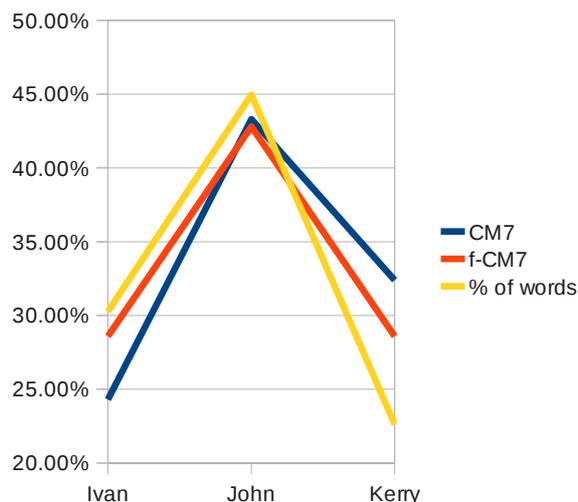


Figure 4. Critical moves and word count

5.3 Construct Linkographs with WordNet of the Two Sessions

Linkography has been criticized for its lack of objectivity in the construction of links, which is primarily based on the discernment and interpretation of the analysts. Also, the process of constructing a linkograph is very time consuming and cognitively demanding [11], making it difficult and impractical to study and compare large data sets. Some analysts utilize the search function to help finding “moves” with similar semantic contents [12]. We proposed an automated method for the construction of linkographs by connecting “moves” using the English lexical database, WordNet [13]. WordNet uses the concept of cognitive synonym (synset) to group words into sets. Words within a synset are connected by meaning. Synsets are also interlinked by means of conceptual-semantic and lexical relations. Synset IDs are assigned to every word. That is words in the same synset share the same synset ID. As a word can have more than one meaning, it can belong to more than one synset and hence several synset IDs. For example the word “surface” contains 10 synset IDs and six of them belong to the noun group.

A Prolog program was written to find the synset IDs of the nouns of each segment. Only nouns were used because adverbs and adjectives are likely produce undesirable links. For example not only all the segments that contain the adverb “on” will be linked but also they will linked to those with the adverb “along”. A Lisp program was written to connect the segments with four or more common synset IDs. This was to account for unwanted association. For example, although WordNet ignore pronouns, “I” in WordNet has the noun meaning of iodine, one, single, unity. We do not want to link segments with only the connection to “I”. Similarly, we do not want to link the segments with the word “here”, “there”, “why”, etc.

5.4 Linkograph Entropy of the Two Sessions

The entropies of the linkographs, produced by this automated method, of the creative session and the pragmatic session are 226 and 187 respectively. We further examine the dynamic variation of entropy during the session. Figures 5 and 6 show the variation of entropy with segment number of the two sessions. We assume the segments were equally distributed and a five minutes window was used to slide across the session. The creative session has 69 segments per five minute window while the pragmatic session has 42 segments per five minute window. The average entropies of this five minute window of the creative and pragmatic sessions are 40.15 and

34.59 respectively. Table 6 summarizes the results of the entropic measurements.

Table 6. Results of the linkograph entropy of both sessions.

A (Creative)	B (Pragmatic)
413 segments	252 segments
Linkograph H=226	Linkograph H=187
69 segments per 5 minutes window	42 segments per 5 minutes window
Average H with 5 minutes window = 40.15	Average H with 5 minutes window = 34.59

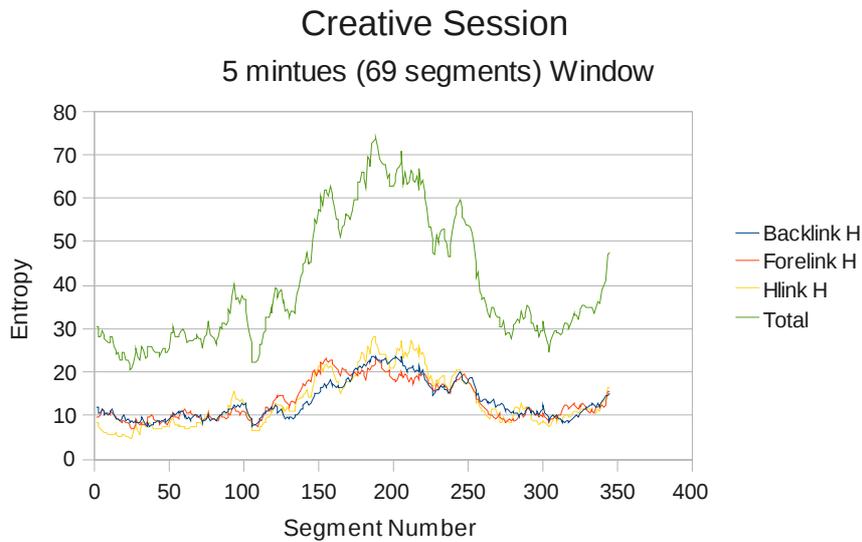


Figure 5. Variation of entropy of the creative session, session A.

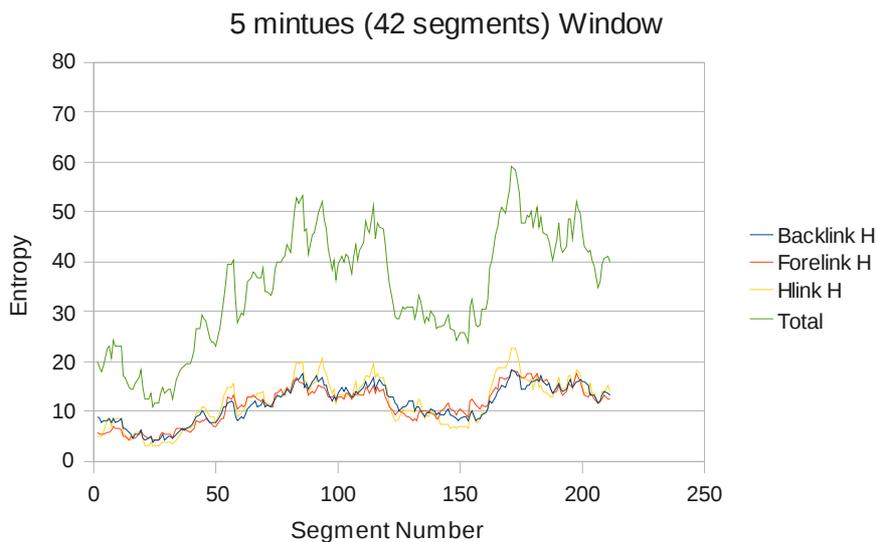


Figure 6. Variation of entropy of the pragmatic session, session B.

According to our hypothesis, a higher entropy indicates a balance of novelty and integratedness, hence creativity. In the creative session, session A, the entropy peaks at segment 189 (entropy = 73.84). The entropy stays above 60 from segment 178 to 222. Segment 178 was the exact time that they started to draw Figure 1(b). The entropy

measured at segment 222, with a 69 segment windows, covered segments until 291. This period, 113 segments, resembles the third episode when they drew Figures 1(b), 1(c) and 1(d). In this period they discussed a few ideas and proposed some shapes such as the idea of a ribbon, a triangular prism with a hole in the middle, a focus to grab attention, ramps to go to the roof, re-discussed inside-outside relationship, and the details of the plan and the section of the proposed building. In this episode they produced 3 sheets of drawing out of the six in total.

6 Conclusions, Discussions and Future Works

The results presented in this paper indicate the potential of using entropic measure as a predictive tool for creative outcome from the design process. Despite the large difference in word counts between the two sessions (4,563 versus 2,309), the text entropies are the same. The creative session has a higher linkograph entropy than the pragmatic session, 226 against 187. The difference is about 20%. Within the sessions, if we measure the entropy with a window of five minutes, the creative session has 345 windows and the pragmatic session has 211 windows. The average entropies of this five minute window of the creative session and the pragmatic sessions are 40.15 and 34.59 respectively. The difference is only about 15%. In the creative session, the period of the highest entropy was the most productive episode. They produced three sheets of drawing (out of six) within this period (about 8 to 9 minutes) and as a consequence of these drawings the session was identified as the most creative session. This measurement matches well with the qualitative analysis.

In this case study the entropy of linkograph of the creative session is higher than the pragmatic session; the entropy variation graphs help to identify the creative/productive period of a session. A larger set of examples are required to test before we can generalize that the measurement of the entropies of the linkographs can be a useful way to study and locate design creativity.

Besides testing the hypothesis on a larger data set, some further analysis that can be done are: examining the forelinks entropy in relation to creativity in this study; comparing the entropy values with the critical move analysis; and exploring the connections of verbs by WordNet. We anticipate these may provide further insight into the relationship between creativity, thoughts and the use of words.

Acknowledgment

Case data was obtained from the CRC for Construction Innovation project titled: Team Collaboration in High Bandwidth Virtual Environments. This research is supported by a grant from the US National Science Foundation, grant no. SBE-0750853.

Reference

- [1] Torres, D.F.M., 2002, Entropy Text Analyzer, 2nd Portuguese National ACM Programming Contest, www2.mat.ua.pt/delfim/delfim/artigos/entropy.pdf.
- [2] Goldschmidt, G., 1992, Criteria for design evaluation: a process-oriented paradigm in Y. E. Kalay (ed) *Evaluating and predicting design performance*, John Wiley & Son Inc., New York, pp 67-79.
- [3] Kan, J.W.T. and Gero, J.S., 2005, Can entropy represent design richness in team designing? in A. Bhatt (ed), *CAADRIA'05 Vol 1*, TVB, New Delhi, pp. 451-457.
- [4] Kan, J.W.T. and Gero, J.S., 2008, Acquiring information from linkography in protocol studies of designers, *Design Studies* **29**, 315-337.
- [5] Kan, J.W.T., Bilda, Z. and Gero, J.S., 2007, Comparing entropy measures of idea links in design protocols: Linkography entropy measurement and analysis of differently conditioned design sessions, *AIEDAM* **21**(4), 367-

377.

- [6] Shannon, C.E., 1948, A mathematical theory of communication, *The Bell System Technical Journal* 27, 397–423.
- [7] Goldschmidt, G., 1995, The designer as a team of one, *Design Studies*, **16** (2) pp 189-209.
- [8] Goldschmidt, G. and Tassa D., 2005, How good are good ideas? Correlates of design creativity, *Design Studies*, **26**(6), 593-611.
- [9] van der Lugt, R., 2003, Relating the quality of the idea generation process to the quality of the resulting design ideas, *International Conference on Engineering Design (ICED)*, Stockholm, Sweden, pp. 19–21.
- [10] Cross, N., Christiaans, H. and Dorst, K., 1996, Introduction: the delft protocols workshop, in Nigel Cross, Henri Christiaans and Kees Dorst (eds), *Analysing Design Activity*, John Wiley & Son, pp. 1–14.
- [11] Kan, W.T., 2008 *Quantitative Methods for Studying Design Protocols*, PhD Thesis, The University of Sydney, Sydney.
- [12] Bilda, Z., 2006, *The Role of Externalizations in Design: Designing in Imagery Versus Sketching*, PhD Thesis, The University of Sydney.
- [13] Fellbaum C. (ed), 1998, *Wordnet: An Electronic Lexical Database*, Bradford Books.