Automatic Analysis of Dreams

Reza Amini

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the Degree of Masters of Arts in Experimental Psychology

School of Psychology

Faculty of Social Sciences

University of Ottawa

Acknowledgment

It gives me great pleasure to thank and acknowledge all those who have contributed to my education; whose results have flourished into this thesis. Firstly, I would like to extend my gratitude to my supervisor Dr. Joseph De Koninck who has freely shared with me his extensive knowledge and experience in the area of academia and more. I appreciate his kindness and support which have manifested in various ways.

Dr. Stan Matwin has advised and guided me in many ways, predominantly in dealing with the artificial intelligence and machine learning aspects of this project. I enjoyed my interactions with him and I'm happy to have learnt so much from this pleasant scholar.

I also thank Catherine Sabourin who helped me with my research in many ways; particularly in using her expertise and offering her precious time in coding hundreds of dream reports.

I am also like to thank all the members of my faculty, and the administrative staff who have treated me with great kindness and respect through out the years. They work and efforts have truly made this experience a smooth experience.

I am also in debt to all the volunteers and honours students who have contributed their time and effort in the laboratory with various tasks, which have incidentally helped my research. I would like to specially thank Christina Wong, Alexandra Duchesne-Perusse, Catherine Joanis-Sirois, Daniel Ouellet, Catherine Joanis-Sirois and Simon Bradshaw.

Finally and most importantly, I would like to attribute my success to all of my friends and family for their emotional support. Without them this experience would not have been the same.

Curriculum Studiorum

Reza Amini was born on March 14, 1980 in Tehran, Iran. He obtained a B.SC. in Computer and Electrical engineering from University of Waterloo in May 2004, and a B.SC. in Psychology from University of Ottawa in May 2009.

Articles

Amini, R., Sabourin, C., & De Koninck, J. (In Press, Prepublication available online, August 27, 2011). Word Associations Improve Automatic Analysis of Dream Emotional Tone (in full). *Consciousness* and *Cognition (http://dx.doi.org/10.1016/j.concog.2011.08.003).*

Matwin, S., Razavi, A., De Koninck, J., **Amini, R.** (2010). Classification of Dreams using Machine Learning. *Proceedings of the ECAI-2010*, 169-174.

Summary of published work

Amini, R., Sabourin, C., & De Koninck, J. (2011). Adverb Modifiers Improve Automatic Scoring of Emotional Tone of Dreams. *Sleep,* 34, A112.

Amini, R., Sabourin, C., & De Koninck, J. (2011). Transformed Attributes Facilitates Automatic Scoring of Emotional Tone of Dreams. *Sleep,* 34, A112.

Amini, R., Sabourin, C., & De Koninck, J. (2010). Semantic Associations Enhance Automatic Analysis of Dream Emotional Valence. *Sleep*, 33, A82.

Amini, R., Ouellet, D., Sabourin, C., & De Koninck, J. (2009). Word Associations Improve Automatic Analysis of Dream Emotional Tone. *Vigilance,* 19, P038.

Bradshaw, S., Ouellet, D., **Amini, R**., & De Koninck, J. (2009). Threat simulation in dreams and the severity of waking threatening experience. *Sleep Medicine*, Vol 10, Supplement 2, 56 (021).

Razavi, A. H., **Amini, R.**, Sabourin, C., Sayyad Shirabad, J., Nadeau, D., Matwin S. & De Koninck, J. (2008). Evaluation and Time Course Representation of the Emotional Tone of Dreams Using Machine Learning and Automatic Text Analyses. *Journal of Sleep Research*, 17, 257.

Razavi, A. H., **Amini R.,** Sabourin, C., Sayyad Shirabad, J., Nadeau, D., Matwin S. & De Koninck J. (2008). Classification of emotional tone of dreams using machine learning and text analyses. *Sleep*, 31, A380-381.

Articles in preparation

Wong, C., Amini, R., De Koninck, J. (In preparation). Automatic detection of gender in dream reports.

Table of Contents

List of Figures

List of Tables

Abstract

In a scientific study of dream content, artificial intelligence has been utilized to automatically score dream content. An initial attempt focused on scoring for emotional tone of dream reports. The contribution of this thesis demonstrates methods by which accuracy of such a system can be improved beyond text-mining. It was hypothesized that data extraction based on psychological processes will provide significant information that would produce an accurate model. In our first article, the significance of words expressed in dream reports, along with their associated words was explored. Extraction and inclusion of these associations provided detailed information that improved automatic scoring of positive and negative affect even though these associations exhibited skewed distribution. The second article demonstrated how normalization of the data was possible and how it could result in a more accurate model. Our last article was able to demonstrate that the model can differentiate between male and female dreams.

Contributions

Article 1

This article was a major extension to my honours thesis. With extended literature review, detailed statistical analysis and a revised discussion, this paper has been published in the journal of Cognition and Consciousness. Dr. Joseph De Koninck provided valuable guidance in this research. He provided great suggestions in terms of refining the methodological approach, writing and submitting this article for publication. He also provided the necessary equipment and encouragement. Dr. Catherine Sabourin assisted this research by scoring dream reports on the positive and negative affect scales; which I used for training a computer model for automatic scoring of emotional tone.

Feedback made by Dr. Pierre Mercier on my honours thesis was also used for the extension of this study to better this article. Dr. Stan Matwin has also provided much valuable suggestions and advice in the area of artificial intelligence.

My unique contribution entailed additional literature review, design, implementation and statistical analysis of the study

Article 2

This article is the result of my work in the context of PSY6042 course (Practicum in Basic Research). This article was completed solely by me. Additional revisions were made upon receiving feedback from Dr. Sylvain Chartier.

Article 3

This article is the result of an honours thesis conducted by Christina Wong, under my supervision and that of Dr. Joseph De Koninck. Christina Wong conducted the literature review on gender differences, experimental design and wrote the article. I introduced and refined the methodological approach of this paper. I implemented and conducted the statistical tasks necessary for machine learning aspect of this paper. To advance this research further, it was decided to code additional dream reports using the Hall and Van de Castle scale (HVdC), and to use these scores in the machine learning tasks. In this initiative, I compiled a set of dreams that were matched for length of dream reports, and number of dream reports. These dream reports were scored on the HVdC scale and entered into DreamSAT by Christina Wong. Catherine Joanis-Sirois also volunteered her time in coding and data entry.

Dr. Monique Lortie-Lussier provided much appreciated insight and guidance with respect to dream content differences between genders.

1.0 Introduction

Dreams have influenced and touched humanity in many ways and our intrigue to understand their meaning has been reflected in our attempts to scientifically study them. Systematic approaches to studying dreams have proven to be difficult, mainly due to the introspective and descriptive nature of dream reports. These experiential reports are qualitative data that are rich in content and cover a vast range of domains. Each dream is unique and for this reason systematic approach to research of dreams has focused on common elements and thus has limited its scope in this way. But this has not been the primary difficulty in this field of research. Dreams have been accessible only through oral accounts as expressed subjectively and selectively by the dreamer. This subjective nature of these qualitative data has been one part of the challenge. In addition, discerning meaning from these reports has been prone to the subjective biases of the evaluator. These two aspects of studying dreams are amongst the many challenges of scientific dream research.

Recently these qualitative dream reports have been quantified using comprehensive scales which give codes and values to different facets of the report. The most comprehensive scale has been developed by Hall and Van de Castle (1966). The Hall and Van de Castle system (HVdC) probes eight dimensions ranging from emotions to characters. These categories aimed to measure common yet global elements. Frequencies of these codes have often been used to compare differences between groups. Management of these codes and their usability in conducting statistical analysis have been greatly facilitated by Dreams AT; a system whereby codes are translated and prepared for statistical analysis through the use of spreadsheets (Domhoff 2003).

Theories of dream production have been used to guide research to focus content extraction and coding of that content on rudimentary features of dreams. There have been many theories that have attempted to explain the arising of dreams. Foremost, dream production has been attributed to the result of cognitive processes (Foulkes, 1985; Cavallero & Foulkes, 1993). Conversely, there has been an increased interest in the emotional component. Hartmann (2007) has proposed that dreams evolve with emotions as their foundation. Support for this notion has come from researchers that have linked dream emotions to the underlying physiology of REM sleep during which vivid dreaming is predominant (Mancia, 2005; Benca, Obermeyer, Shelton, Droster, & Kalin, 2000; Dang-Vu, Schabus, Desseilles, Schwarts & Maquet, 2007). More specifically, these observations seem to be consistent with the prevalence of negative emotions in REM dreams (Roussy, Raymond & De Koninck, 2000). In order to test these theories in light of the prevalence of emotions in dreams, it would be necessary to have means for quantifying the emotional content of dream reports and evaluate the degree of a particular emotional expression in dream reports.

On the emotional scales of the HVdC system, Anger, Apprehension, Happiness, Sadness and Confusion can be scored for their presence or absence. In order to further investigate these emotions in greater detail, it would be necessary to quantify the degree to which these emotions are expressed. In this task might, reliability of the scales of emotional degree, as scored by a human judge, may be compromised due to variability within the judge as well as his/her subjective biases. Therefore, a rigorous and systematically reliable means for scoring this would be useful. Early attempts by Nadeau, Sabourin, De Koninck, Matwin, & Turney, (2006) have established the practicality of this by trying to set up a baseline for reliability of the overall expression of positive and negative affect on a 4 level scale (0=none, 3=high). They have showed that the degree of agreement between trained professional dream researchers was 81% (MSE = .19) on the negative affect scale and 58% (MSE = .54) for the positive scale. These results showed that the suspected variability in scoring emotional valence was in fact present. As part of their research, an artificial intelligence text-mining model was used to automatically score emotional tone on the same scale. The model was trained with a set of dreams as scored by a human judge and tested against a hold out set which proved to have an accuracy of 48% (MSE = .608). Despite the lower than inter-human judge accuracy, this tool had the advantage of reliability and consistency. In an ensuing study, Razavi et al. (2008) implemented means for measuring the emotional fluctuations within dreams to improve their automatic scoring of the overall negative affect of dream reports as per Hartmann's (2007) theory of emotionally driven dream. In their study, inter-word relations were established using proximal co-occurrence as suggested by Barcaro et al. (2005) who established that words may be linked to other sources. Their study showed an improvement of 59% in accuracy on the 4 level negative-affect scale. Recently Matwin, Razavi, De Koninck & Amini (2010) took one step further by incorporating co-occurrence vectors and advance machine learning techniques to achieve an agreement of 64% on the same scale.

It is evident that this line of research has demonstrated that the overall negative tone of a dream report can reliably be scored automatically using artificial intelligence. With further advancements, one may anticipate further improvements in such a tool, and also in its application toward scoring specific emotions. To this end, the findings of the first two of there three articles show how the accuracy of the automatic scoring of emotional tone of dreams can be improved. The first illustrates how word associations, as established using external sources, may replace the proximal co-occurrence word relationships that were previously built using the same dream reports used for evaluation. The second article demonstrates the necessity of data normalization for building a more accurate model. The outline of these articles follows.

Matwin et al (2010) used proximal word co-occurrence vectors (PWCV) to enhance the text-mining aspect of their latest model to automatically score emotional tone of dreams. This was an excellent attempt to exhibit the inter connectedness of words. These connections showed that meaning is not just expressed in discrete words, but rather through the connections that they imply. These connections were established using the text from all dream reports. It can be argued that the reuse of dream reports for establishing PWCV and for evaluation may be in confounding variable. While the model was tested on a hold out set, the network of word associations, established in the preprocessing stage, held information that pertained directly to the hold out set. In order to resolve this potential confound, it was deemed an important investigation to establish word-associations from a source outside of the bank of dream.

Furthermore, PWCV seem to produce word associations that resembled associative memory that is developed through coincidental semantic and episodic factors (Silberman, Bentin, & Miikkulainen, 2007), which are syntagmatically structured in childhood. Hence, it may be argued that this type of associations may not be appropriate for adult dreamers. As suggested by Gomes (1995), a paradigmatically structured association better reflects the semantic word associations in adults. Therefore, semantic connections established on word definitions might be more appropriate in this case than the PWCV.

In order to advance this line of research it is necessary to resolve these two confounding factors by establishing word relations using an external source. Those relationships can be established using paradigmatic approach as opposed to proximal coincidence to appropriate the word relations to adult. In this endeavour, word definitions as found in dictionaries and encyclopedias, can be used in order to establish word associations which reflect the common knowledge. These associations can be made between the words expressed in definitions to the words they define. Concurrently, the PWCV is avoided and word relations better reflect a paradigmatic representation. Our first article entitled *"Word Associations Contribute to Machine Learning in Automatic Scoring of Degree of Emotional Tones in Dream Reports"*, addressed these two factors. It was hypothesized that emotional valence (positive and negative) could be reliability predicted with the use of word associations (established from external sources) as a substitute for PWCV (established using dreams under evaluation).

Word associations established based on a whole dictionary and encyclopedia proved challenging, primarily in one respect. The large number of words, and the greater number of possible word-word associations, poses a problem of dimensionality. The number of associations was overwhelming to organize with the use of computers. But one inherent problem was the numerous associations that common words had with all of the other words. As such, associations exhibit a skewed distribution. There were few strong relationships and many others which monotonically depreciate exponentially. To resolve this problem, it was suggested to normalize or re-distribute these associations using commonly used mathematical functions. This would facilitate the machine learning process and classification and thereby result in an improvement in performance.

An initiative was taken to address the data distribution issue associated with word associations in our second article entitled *"Data Normalization Assists Automatic Scoring of Emotional Tone of Dreams"*. The problem with data normalization was mainly due to the fact that there were too many variables to be handled individually; about 130,000. Therefore, this process had to be addressed in an automatic manner in order for this initiative to be practical.

In automatic scoring of emotional tone of dreams, data is extracted from dream reports and good predictors are selected in a process called attribute selection. In this routine, attributes that uniquely account for a substantial percentage of the variance in emotional tone are selected and the rest is discarded. Under this premise, one may include normalized versions of the original data and expect that they would be selected only in cases where they are better than their original data for the model. As such it was hypothesized that the accuracy of automatic scoring of positive and negative valence of dream reports would improve with inclusion of transformed data using square-root, square and log functions (commonly used to normalize skewed data).

Automatic scoring of emotional tone of dreams can be advanced to further improve the performance of the models. Alternatively, it can also be expanded to score other aspects of dream reports. In fact, by scoring other aspects one may find validation for the capabilities of the automatic scoring system. One potential aspect that may be explored would be gender differences in dream content. Outside of dream research, automatic techniques have been effectively used to determine the gender of an author using excerpts from their books (Koppel, Argamon, & Shimoni, 2002), suggesting gender differences in expression. Implementation of an automatic scoring of dreamers' gender, could serve as means for validation of the automatic scoring system and perhaps provide evidence for gender specific classification.

Gender differences in dream content have also been noted. Some researchers attributed this phenomenon to biological predisposition (Van de castle, 1966) while others, attribute it to social factors (Lortie-Lussier, Schwab, and De Koninck, 1985; Lortie-Lussier et al., 1992). In either case, gender differences in dream content do exist regardless of their source, and such knowledge can be useful for guiding the future of dream content analysis. One implication would be that automatic analysis of emotional tone of dream reports might benefit from gender specific analysis. To warrant such an endeavour it must be shown that the artificial intelligence behind the automatic analysis can differentiate between male and female dreams. Such evidence would be clear indication that there are measurable differences in content, which may be a confounding factor when studying male and female dreams together.

In a joint study, automatic scoring of dream reports was expanded to predict dreamer's gender. As a benchmark, human judges were asked to read dreams and predict the dreamers' gender to their best of abilities. This performance at predicting the dreamer's gender was used as supporting evidence for differences in dream content differences. But primarily, the human judges' accuracy was used as an indication of how well a computer model may perform. In the last article, entitled *"Automatic detection of gender in dream reports"* it was hypothesized that there are differences in dream content between males and females, and furthermore the syntax of dream reports could be used by a machine learning algorithm to predict the dreamer's gender at an accuracy close to that of human judges.

2. Methods and Results: Research Articles

2.1 Article 1

WORD ASSOCIATIONS CONTRIBUTE TO MACHINE LEARNING IN AUTOMATIC SCORING OF DEGREE OF EMOTIONAL TONES IN DREAM REPORTS

Reza Amini, Catherine Sabourin PhD, Joseph De Koninck PhD

School of Psychology, University of Ottawa, Ottawa, Ontario, Canada.

In press: Cognition and Consciousness, 2011.

Abstract

Scientific study of dreams requires the most objective methods to reliably analyze dream content. In this context, artificial intelligence should prove useful for an automatic and non subjective scoring technique. Past research has utilized word search and emotional affiliation methods, to model and automatically match human judges' scoring of dream report's negative emotional tone. The current study added word associations to improve the model's accuracy. Word associations were established using words' frequency of co-occur with their defining words as found in a dictionary and an encyclopedia. It was hypothesized that this addition would facilitate the machine learning model and would improve its predictability beyond those of previous models. With a sample of 458 dreams, this model demonstrated an improvement in accuracy from 59% to 63% (kappa = .485) on the negative emotional tone scale, and for the first time reached an accuracy of 77% (kappa = .520) on the positive scale.

Keywords
Dream content, Dream emotions, Emotional tone, Artificial Intelligence, Automatic analysis, Cognition, Word association, Emotion progression,

Introduction

Scientific research on dreaming and dream content has strived to establish objective methods of collecting dreams and analyzing them. In order to limit the distortion and the loss of recall of dreaming activity, subjects are asked to immediately report their dreams upon awakening in laboratory or at home. The reports are then coded using a number of scales that have been developed (Winget and Kramer, 1979). The most comprehensive system was developed by Hall and Van de Castle (1966) which has eight dimensions ranging from characters to emotions. Domhoff (1996) has improved its application notably with DreamSAT which utilizes spreadsheets as means for data entry and statistical analyses (Domhoff, 2003a).

Such a scoring system relies on human judges' reliability and inter-judge agreements as a requirement for scientific studies (for example, Barcaro, Cavallero & Navona, 2005). As such, these studies may benefit from an automatic scoring technique. Our current research program attempts to develop an automatic scoring system that ultimately can match human judges' and the dreamer's own scoring. As a first step, we have focused on dream emotions. While it is well established that dream production is primarily a cognitive process (Foulkes, 1985; Cavallero & Foulkes, 1993), there is increased interest in the emotional component of dreams. Hartmann (2007) has gone so far as to propose that dream construction starts from central emotions. There are also some physiological studies that attempt to implicate emotions in REM sleep (Mancia, 2005; Benca, Obermeyer, Shelton, Droster, & Kalin, 2000; Dang-Vu, Schabus, Desseilles, Schwarts & Maquet, 2007). These observations are consistent with the typical prevalence of negative emotions in REM dreams (Roussy, Raymond & De Koninck, 2000).

Emotions in dreams can be evaluated by human judges who read the transcript of dreams or by the dreamer. Studies have shown discrepancies between human judges and dreamer in estimating levels of emotions (Fosse, Stickgold & Hobson, 2001). For example, dreamers attribute more emotions to their dreams than human judges. The Hall and Van de Castle system has five subscales: Happiness, Apprehension, Anger, Sadness and Confusion, which have been derived after extensive research (reviewed by Domhoff, 2003b). Given the complexity of these dimensions, however, we have chosen to start with two four-level scales to score the global positive and negative emotional tone of dreams in an attempt to model the human judge's scoring, using machine learning. These scores are global measures of the degree of an emotion within a dream which may rely on the dynamics of emotions within the narrative

In a first study, experienced dream researchers scored the emotional tone of 100 dreams. With disagreements rarely more than one level apart, it was found that the degree of agreement for negative sentiment of dream reports was 81% (MSE = .19) and for positive sentiment it was 58% (MSE = .54) (Nadeau, Sabourin, De Koninck, Matwin, & Turney, 2006). Therefore, further work concentrated on the negative scale. An application called Linguistic Inquiry and Word Count (LIWC) was first used to identify words pertaining to expression of positive and negative emotions in each dream report. The frequency of such words was used to generate additional attributes using mathematical functions; log ratio, square and square root. These attributes were used to train a linear regression model to predict negative affect score of a human judge on a subset of all dream reports, and tested against the remaining dream reports. The model proved to be efficient with an accuracy of 48% (MSE = .608) (Nadeau et al., 2006). The use of individual words as attributes in modeling a dream report was in line with other dream content analysis research (Bulkeley & Domhoff, 2010). In a second study, Razavi et al. (2008) recognized that a more accurate model could be implemented by including patterns of emotional fluctuation; deemed a good source of attributes as per Hartmann’s (2007) theory of emotionally driven dream. As such, attributes of emotional fluctuations, such as frequency of rises/falls of emotions, length of rise/falls, and maximum and minimum, were quantified and used in the model building. These attributes were recognized as Emotion Progression attributes. Finally in the same study, it was suspected that words may be linked to other sources (Barcaro et al., 2005) and to other words through proximal co-occurrence. Hence, the number of words separating any two words was used as an approximation of their semantic closeness. In this manner, a new set of attributes was generated and used in this study. In conjunction with all these extracted attributes, the dreamer's own report of Joy, Fear and Anxiety, and Hall and Van de Castle's dimensions of Anger, Apprehension, Happiness, Sadness and Confusion were included. Where Hall and Van de Castle recognized the presence or absence of these dimensions, here the dreamers were asked to report the emotional intensities on a 4 level scale. In this experiment, the model produced an accuracy of 59% for negative sentiment. Matwin, Razavi, De Koninck & Amini (2010) took one step further by incorporating co-occurrence vectors and advance machine learning techniques to achieve an agreement of 64% on the negative scale.

In the present study, we improve upon Razavi et al.'s (2008) model by replacing their word co-occurrence vectors (established using the dream reports under analysis) with word associations (established from external source). This is based on the notion that associative memory is used by dreamers when they report, and by human judges when they read and score for emotional tone. In this process it is assumed that there is communication of meaning indirectly through written words. This shared meaning exhibits itself in part in the network of word associations. Silberman, Bentin, & Miikkulainen (2007) suggest that these associations form over our life time, as semantic and episodic factors coincide. These associations can be observed to be syntagmatically structured in childhood and a paradigmatically structured in adulthood (Gomes, 1995). In analyzing dream content of adults, the network of word associations can be built using words expressed in dictionary and encyclopedia definitions, as a whole, in order to better approximate paradigmatic word association. Alternatively, sequential proximity of words in definitions would better approximate syntagmatical structure which would benefit analysis of dream content for children. The frequency of word-coincidences can be a good basis for word association, although high frequencies do not necessarily mean high word association as suggested in the Search of Associative Memory (SAM) Model of semantic memory (Nelson & McEvoy, 2000; Raaijmakers & Schiffrin, 1981). As such, in this study, we normalized these frequencies starting with the highly frequent words first.

Incorporation of word-association inherently demands a different data structure than those of the past. Recently, word search (Bulkeley et al., 2010), word-strings (Domhoff, & Schneider, 2008) and bag of words (Nadeau et al. 2006) provided means to identify and count specific occurrences. With the assumption that there is a hierarchical structure in meaning (Murphy & Lassaline, 1997), these frequencies were then used for grouping words to form discrete categories (Bulkeley, 2009) or fuzzy categorical membership (Frantova & Bergler, 2009) which were used to measure specific aspects: personality, emotions, etc (Hall & Van de Castle, 1966). These aspects are quantified by means of search, which have the strength of objectivity, but suffer from narrow scope of observation due to discretization. It is proposed that word-association can open up this scope for a greater perspective on dream content. One dream report, as represented by a network of associations, can be compared to another's for similarities and differences using machine learning algorithms. In the former paradigm, one is required to search for specific words in order to quantify and compare frequencies. That approach is limited by its narrow scope; by focusing on a few variables, one has to ignore other potential variables.

In the proposed paradigm, differences in the network of associations can be identified using machine learning, relinquishing the need to focus on specific items.

Therefore, it was hypothesized that, with the greater scope of and access to implicit information provided by word associations, the accuracy in automatic scoring of the emotional tone of dreams as assessed by human judges will improve. More specifically, it was predicted that the inclusion of associated words would increase the percentage of agreement between machine and human judges.

Methods

Participants

Dream reports were submitted by participants to the Normative Study of Dreams of Canadians, conducted at the University of Ottawa. Upon signing a consent form, participants were instructed to keep a journal of their daily events, and their dreams. The task terminated when 10 consecutive daily event questionnaires were completed or 2-4 dreams had been reported. A subset of 458 English dreams met the 50-800 word count limit, and were used for this study. These dreams were reported by 172 females and 51 males.

Classifier: Human Judge Scores

A classifier is the variable that the machine learns to predict, and in this case, it is the emotional tone of the dream report according to a human judge. Human judges were given a package of 458 dreams and were asked to read each dream report and give a score for positive and negative affect on a scale of 0-3, representing Not at all, A little, Moderate and A lot. This task was given to two experienced judges with good reliability in scoring dreams; particularly with the Hall and Van de Castle system. One of them (author C.S.) was one of the judges who scored the emotional tone of dreams in the previous studies (Nadeau et al., 2006; Razavi et al., 2008; Matwin et al, 2010). The inter-judge reliability on the negative scale was 66% (kappa =

.529) and 76% (kappa = .524) on the positive scale. Given the improvement in reliability of the positive scale, it was included in the machine learning. In training the model the scores used were those of C.S.

Design and Procedures

In the Canadian Normative Study, participants were instructed to write down his/her dream(s) upon awakening in the morning. Following each dream report, the dreamer was instructed to fill out a questionnaire which includes a scale of 1-4 to score the degree to which they experienced the following emotions: happiness, apprehension, anger, sadness, confusion, joy, fear and anxiety. The first five scales are that of the Hall and Van de Castle system (1966) originally scored for presence or absence. The last three scales are part of the Normative Study, and will be assessed using a much larger sample. These eight items constituted the first class of attributes called Dreamer's Emotions. The dream reports were typed into a text document by volunteers without any modification and used to extract additional classes of attributes.

Preprocessing: A word-association matrix was constructed as a source for an approximate probability of semantic closeness between any two English words. This matrix was built using all of the words used in the dream corpus and the words in their definition as found in two independent sources; namely Wikipedia.org (WIKI) and WordReference.com (WR). Each word was assigned a unique identification number which corresponded to the word's frequency of use.

An independent square matrix was constructed for WIKI and WR. Each cell of the matrix was designed to hold a numeric count of word-word concurrence. Occurrences of each word in a definition incremented the corresponding cell in the matrix to account for word and defining word co-occurrence. These counts were normalized to produce a probability estimate of word association. The most frequently used word was normalized first to establish its weak semantic association to other words (Nelson et al., 2000; Raaijmakers et al., 1981). Subsequent normalizations were conducted using the next most frequently used word and the probabilities already accounted for by previous normalizations.

Using a word's identification number as vector an index, one can easily compile vectors that represent the probabilistic association of a word to all other words. This form of representation lends itself to mathematical operations. Figure 1 illustrates a two dimensional representation of how one might find closeness of two vectors (words). The dot product of two vectors is a good numerical indicator of degree of closeness between the two words they represent. A dot product of 1 means that the vectors are the same, and 0 would mean that they have nothing in common.

Figure 1. Word Vector Relationship.

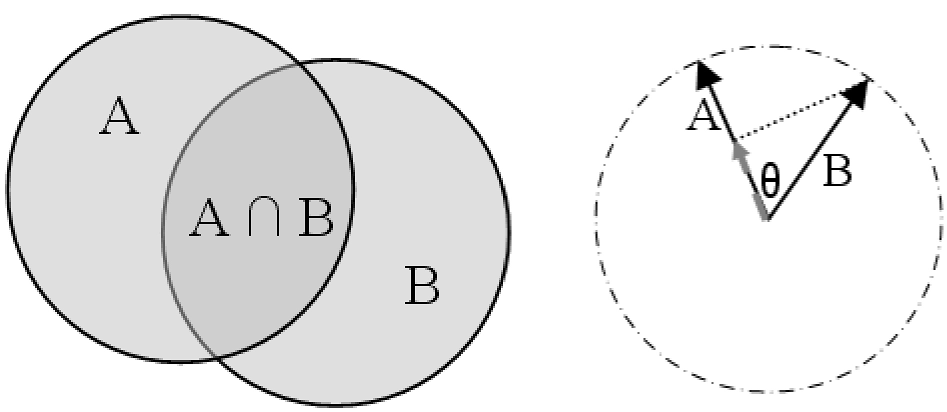

Figure 1. Word Vector Relationship. The Venn diagram shows associations of words to words A, and B as sets with intersection A∩B which can be represented as counts of common elements. Alternatively words can be represented as vectors whose dot product represents a degree of closeness. The vectors are normalized ($|A|=|B|=1$) to accommodate the relative measure of commonality; $A \cdot B = \cos(\theta)$.

Additionally, a collection of vectors can be summed in order to quantify an abstract symbol. In this way, LIWC was used to identify positive and negative words in order to create two Affect Vectors which represent the abstract positive and negative affect. Affect Vectors for each source were constructed, and normalized independently. To test the face validity of this process, word vectors of a sample of words were extracted and their dot products against the Affect Vectors were calculated. “Angry” and “sad” were selected from LIWC’s list of negative words and “happy and “joy” were selected from the list of positive words. “Shouted”, “guessed”, “completed” and “glow” were included as foreign words, since they were not recognized by LIWC as positive nor negative. “Shouted” and “guessed” are expected to be associated to the word “angry” and thus associated to Negative Affect. “Completed” and “glow” are expected to be associated to “joy” and thus associated to Positive Affect.

Attribute Extraction: For each dream, the Word Vectors of its words were summed in order to represent the whole content of each dream with one vector (Dream Vector). The dot product of this vector with the Affect Vector yielded a numeric value representative of how closely the whole dream is associated with the affect. The set of dot products and the raw Dream Vectors were recognized as the Word Association class of attributes. Word count was also added in order to help account for the length of the dream report.

The dream report’s array of Word Vectors is the source of data for the next stage of data extraction; namely Emotion Progression attributes. The dot product of each Word Vector of a dream with an Affect Vector results in a sequence of fluctuations of associated affect. The numerical array is treated to a digital signal filter in order to remove high frequency noise (error) in order to retrieve a smooth function. The resulting array represents an estimate of the progression of the dream’s emotional progression. Figure 2 shows the fluctuation of the positive affect for a particular dream report. For each increase or decrease in these arrays the following attributes are calculated and stored: frequency, average, maximum, minimum and standard deviation (Razavi et al., 2008). This set of data is recognized as Emotional Progression attributes.

Figure 2. Positive Affect Progression in a Dream Report.

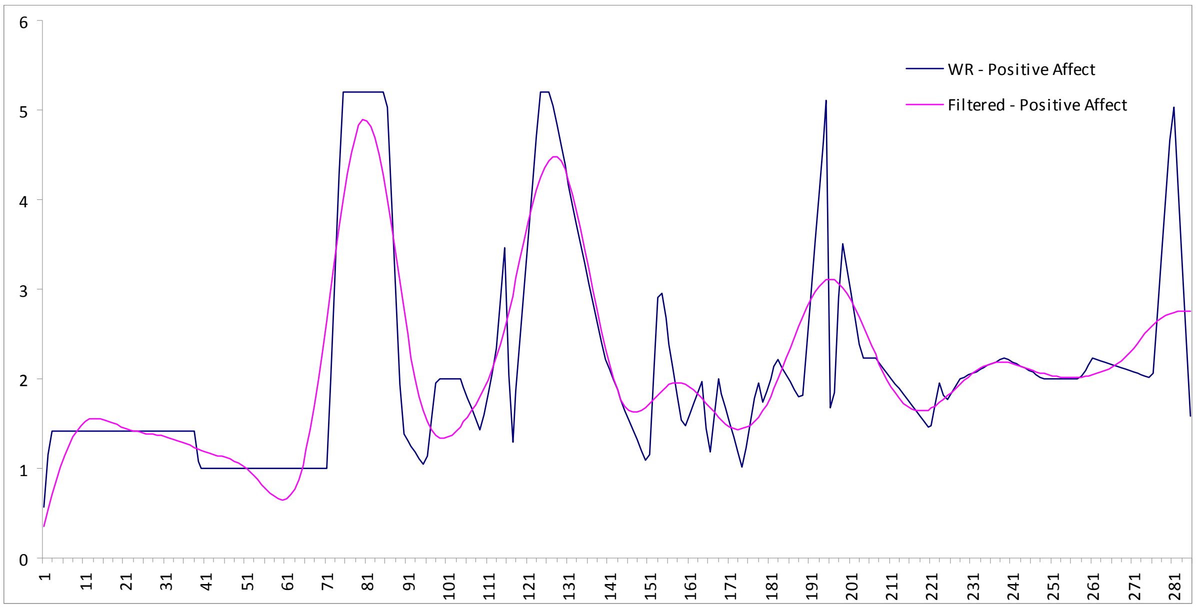

Figure 2. Positive Affect Progression in a Dream Report. The following graph shows the raw and filtered positive affect of a sample dream report. LIWC recognized words are in bold. The dream report reads: "*My **friend** 's wedding. Not my closest **friend** and she has no plans to marry as of yet. We were at her house. But not her real house. And it was really old and dumpy they were renovating at the same time. But they had this ballroom that we were going to use for the wedding. We hang around for a bit and do nothing but then all of a sudden my **friend** decides its time to get started. I am the maid of **honor**. She starts getting **ready** but she is put in charge not me. Then all of a sudden a huge shipment of tiles for her floors arrives. My dad is there and asks her where all the tiles go and he starts tiling. But my **friend** does not remember. I know but I can not speak. We get dressed and I have to wear an ugly fuchsia dress. We have no wedding ceremony only a reception dinner in her ballroom which is unfinished. There is no floor walls are cracked and cobwebs are everywhere. She tells me I have to sit on her right at this long long table. She is at the head of the table but I actually sit on her left because I decide that is what she actually meant she does not seem bothered by this. I try to start a toast but I still can not speak*"

Attribute Selection: Attribute selection is the process where a small sample of attributes is selected from a pool of attributes as good predictors. Weka 3.5.8's CfsSubsetEval selection algorithm selects attributes as they compete in their pool of attributes. Ideally one would include all possible attributes in one round of attribute selection. Due to the number attributes that need to be processed along with system limitations, this is not currently practical. Therefore a method called selection in parts is adopted, where selection is applied to smaller blocks, and the selected attributes of each block is combined for a collective reselection. Through this process, it is possible to have redundant and noisy attributes make their way to the final selection stage as a function of chance, thus reducing the performance of the model. Thus, Dreamer's Emotions, Wikipedia-Progression, Word Reference-Progression, Wikipedia-Word Association and Word Reference-Word Association were subjected to attribute selection independently with block size of 1000 attributes. The selected attributes were combined and reselected as a whole.

Machine Training and Testing: The Simple Logistic Regression model was used as the core of machine learning. This algorithm generated a regression for each level of the affect scale. The training data consisted of 66% random sample of the 458 dreams and the remaining sample was used for testing. The performance of the regression was measured by inter-rater (machine-human) reliability score of percentage of agreement, standard error and a Cohen's kappa.

In order to avoid favouring the best or worst case scenario (as selected by chance), the process was subjected to a 10 fold cross-validation. Through repetition and use of different random seeds, the performance was normalized to represent the normal nature of the learning algorithm.

Results

For the face validity test, the associations of a sample of words to positive and negative Affect Vector were calculated using the described dot product method. It is clear from the charted results (Figure B.3), that the words we suspected to be affiliated to positive affect and negative affect, do indeed meet expectations. A thorough validation is warranted for future research.

Figure 3. Face Validity Test Affect Vector.

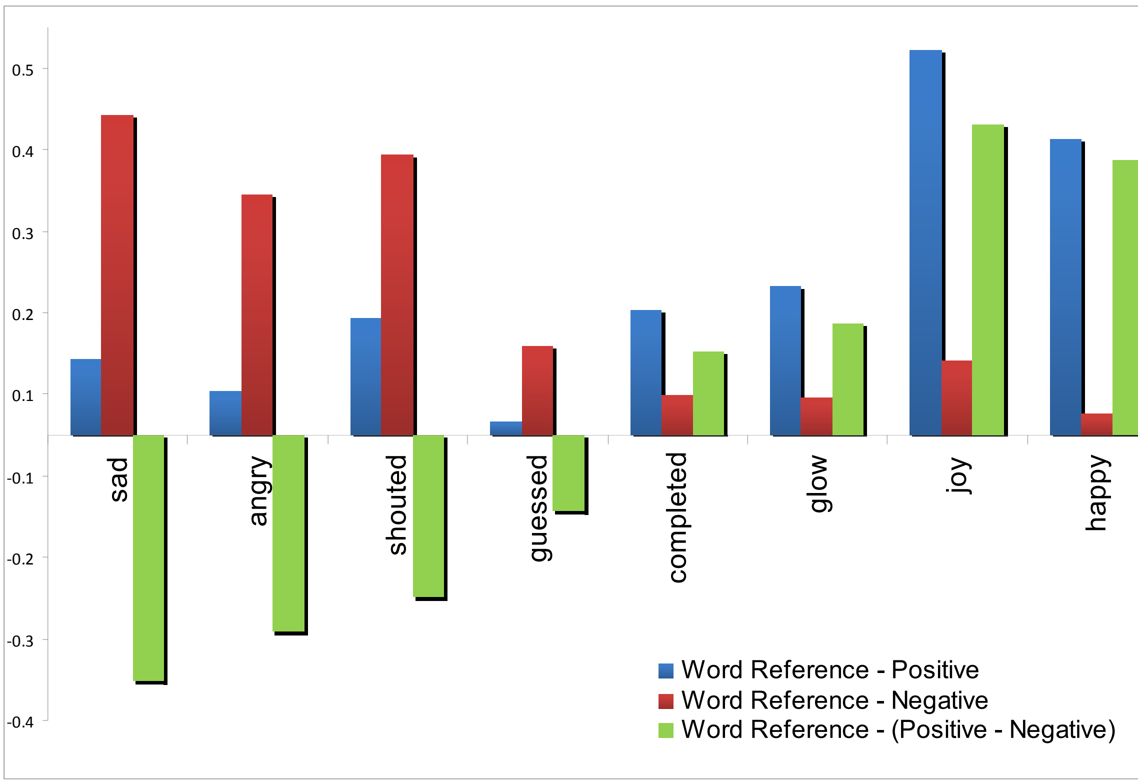

Figure 3. Face Validity Test Affect Vector. The dot product of the Positive and Negative Affect Vector and Word Vectors are charted for a sample of words. The Word and Affect Vectors were built using WordReference.com

Machine performance results showed that the machine-human agreement score was 62.5% (kappa = .485, *MSE* = .388) on the negative affect scale. On the positive scale, the agreement was 76.9% (kappa = .52, *MSE* = .317). These scores were significantly higher than the chance agreement of 25%, and according to Cohen's Kappa scores, they were in *moderate agreement*. The improvement in performance on the negative scale compared to the previous model was 3.5% and agreement of the model for positive affect is entirely new. Finally, inspection of both models revealed that word count was not a good predictor, and therefore did not account for any increments in performance.

Discussion

This paper demonstrated that positive and negative affect could be represented using a set of words as extracted from LIWC. Furthermore, with artificial intelligence as means for learning and handling large volumes of data, it is observed that word associations enhance automatic scoring of dream emotional tone. The contribution rests in enriching the dataset and in describing dream reports beyond a set of discrete words. As a result it has allowed for a more accurate emotional progression whose characteristics were better predictors of the overall emotional tone of dream reports as evaluated by human judges. Integration of word associations opens the door for further application of automatic scoring of other dream scales.

As the next step, one may start coding for more complex emotional scales. One important step will be to score for the dreamer's own rating of personal experience of emotions used by the HVdC system and perhaps that of fear and threat as proposed by the Threat Simulation Theory (Revonsuo, 2000). This will have the advantage of refocusing research on the dreamer's own experience. Subsequently, the dreamer may be asked to give his/her score of the emotional progression within a dream, in order to validate the machine's score of emotional progression. Such an approach will be of interest for the study of emotional progression in nightmares and perhaps, in the case of REM dreams, its electrophysiological and brain imaging correlates.

Further progress will come with the development of this model to predict the scoring on other scales of the Hall and Van de Castle system. Facilitated by the large database of scored dreams on the DreamBank.net (Domhoff & Schneider, 2008), this initiative would also serve to cross validate the models. The volume of dream reports of both male and females will help correct the current imbalance in our sample and provide the possibility for gender specific models.

The described method for incorporation of word association can be a way to broaden the perspective and approach to dream content analysis. Whereas before one was restricted to word searches and questionnaires, today one can automatically search for abstract ideas (positive and negative affects were exhibited in this paper). Nevertheless this method has some challenges and limitations. Firstly, sources (word definitions) used to establish word associations must be representative of the common knowledge that the dreamers have. Using definitions relevant to specific demographic (age group, gender, cultural ethnicity, etc) might help to reduce this error. Alternatively, researchers can study each group of distinct demographic separately. Another obstacle related to these associations, is their skewed distribution that is likely to hinder the machine learning process. Hence, data normalization would benefit the process and improve accuracy. Finally, the group of words that describe an abstract idea (positive, negative affect, etc) must be agreed upon by researchers from different laboratories or established via experimentation. Another challenge that may yet require a creative solution has to do with visual representation. Word associations extend over thousands of words and therefore it becomes difficult to list, communicate and find the exact source of each association. This is a natural phenomenon that arises when one takes into account the depth and breadth of descriptive data.

Perhaps, in the future we may find a convenient method to graphically represent these associations and their inter-connection.

Emergence of a new and systematic tool for analysis of qualitative data has been identified with a growing accuracy in detecting emotions. This may be expanded to explore and code for other dimensions in dreams. With such a tool, dream research may uncover areas and features of dream reports that previously seemed impractical. Ultimately, it may be possible to compare vast numbers of dreams and inspect them graphically.

References

Barcaro, U., Cavallero, C., & Navona, C. (2005). A Method for Recognizing and Describing the Links Among Dream Sources. *Dreaming*, 15(4), 271-287.

Benca, R. M., Obermeyer, W. H., Shelton, S. E., Droster, J. & Kalin, N. H. (2000). Effects of amygdala lesions on sleep in rhesus monkeys. *Brain Research, 879*, 130-138.

Bulkeley, K. (2009). Seeking patterns in dream content: A systematic approach to word searches Consciousness and Cognition, 18(4), 905-916.

Bulkeley, K., & Domhoff, G. W. (2010). Detecting Meaning in Dream Reports: An Extension of a Word Search Approach. *Dreaming*, 20(2), 77-95.

Cavallero, C. & Foulkes, D. (1993). *Dreams as Cognitions*. New York: Harvester Wheatsheaf.

Dang-Vu, T. T., Schabus, M., Desseilles, M., Schwarts, S., & Maquet, P. (2007). Neuroimaging of REM sleep dreaming. In Barrett, D., & McNamara, P. (Eds.), *The new science of dreaming vol 1* (pp. 95-114). London: Praeger.

Domhoff, G. W., & Schneider, A. (2008). Studying dream content using the archive and search engine on DreamBank.net. *Consciousness and Cognition*, 17(4), 1238-1247.

Domhoff, G.W. (1996). *Finding meaning in dreams: A quantitative approach*. New-York: Plenum Publishing.

Domhoff, G. W. (2003a). *A new resource for content analysis*. (pp. 95-105). Washington, DC, US: American Psychological Association.

Domhoff, G.W. (2003b). *The Scientific Study of Dreams*. Washington, DC, US: American Psychological Association.

Fosse, M. J., Stickgold, R. J., & Hobson, J. A. (2001). The mind in REM sleep: Reports of emotional experience. *Sleep, 24,* 947-955.

Foulkes, D. (1985). *Dreaming: A Cognitive-Psychological Analysis*. Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Frantova, E., & Bergler, S., (2009). Automatic Emotion Annotation of Dream Diaries. In: *Analyzing Social Media to Represent Collective Knowledge* Workshop at K-CAP 2009, The Fifth International Conference on Knowledge Capture.

Gomes, H. T. (1995). An examination of syntagmatic and paradigmatic associations using behavioral and electrophysiological measures of priming tasks. (Doctoral Dissertation City University of New York, 1995). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 55*, 5103.

Hall, C., & Van de Castle, R. (1966). *The Content Analysis of Dreams*. New York: Appleton-Century-Croft.

Hartmann, E. (2007). The nature and functions of dreaming. In D. Barrett & P. McNamara (Ed.), *The new science of dreaming: Volume 3* (pp. 171-192). Westport, CT, US: Praeger Publishers.

Mancia, M. (2005). The dream between neuroscience and psychoanalysis. *Schweizer Archiv fur Neurologie und Psychiatrie. 156*(8), 471-479.

Matwin, S., Razavi, A., De Koninck, J., & Amini, R. (2010). Classification of Dreams using Machine Learning. *Proceedings of the ECAI 2010*, 169-174.

Murphy, G., & Lassaline, M. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 93-131). Cambridge : MIT Press.

Nadeau, D., Sabourin, C., De Koninck, J., Matwin, S., & Turney, P.D. (2006, July). *Automatic dream sentiment analysis*. Poster session presented at the Twenty-First National Conference on Artificial Intelligence (AAAI-06), Boston, Massachusetts.

Nelson, D. L., & McEvoy, C. L. (2000). What is this thing called frequency? *Memory & Cognition, 28*(4), 509-522.

Raaijmakers, J. G. W., & Schiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *8* (2), 98-134.

Razavi, A. H., Amini, R., Sabourin, C., Sayyad Shirabad, J., Nadeau, D., Matwin, S., & De Koninck, J. (2008). Evaluation and Time Course Representation of the Emotional Tone of Dreams Using Machine Learning and Automatic Text Analyses. *Journal of Sleep Research*, *17*, 257.

Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and Brain Sciences, 23,* 877–901.Roussy, F., Raymond, I. & De Koninck, J. (2000). Affect in REM dreams: Exploration of a time-of-night effect. *Sleep, 23*, A174-A175.

Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: An empirically-based computational model. *Cognitive Science: A Multidisciplinary Journal, 31*(4), 645-671.

Winget, C., & Kramer, M. (1979). *Dimensions of dreams.* Gainesville: University of Florida Press.

2.2 Article 2

DATA NORMALIZATION ASSISTS AUTOMATIC SCORING OF EMOTIONAL TONE OF DREAMS

Reza Amini

School of Psychology, University of Ottawa, Ottawa, Ontario, Canada.

Abstract

The goal of this research program was to improve the automatic scoring system of the emotional tone of dreams by incorporating the mathematical transforms of all attributes to correct for possible skewed data distribution. A human judge scored 457 dream reports on 1 to 4 negative and positive affect scales. Our computer model was trained with 66% of 457 dream reports and tested on the remaining dreams for both scales. In the attribute extraction stage of the machine learning, the data was mathematically transformed with log, square and square-root functions. Cohen's kappa coefficient for computer-human agreement was calculated and compared to our previous results. On the positive scale, the machine human agreement remained at 78.7% (kappa 0.54) and improved from 67.3% (kappa 0.55) to 69.2% (kappa 0.57) on the negative scale. This study shows that transformed data is beneficial in the automatic scoring for negative affect.

Introduction

In studying dreams, participants give an oral account of their dreams upon awakening at home or in the laboratory. These reports are then transcribed and scored on various dimensions, and thus objectified for statistical analysis. Scoring dream reports require human intervention which is laborious and often a source for error. In an attempt to automate this process an initial attempt is made to score emotional tone as a global measure of dream reports through employment of machine learning. Machine learning entails extractions of syntax based attributes from dream reports, and model training with human assistance. While still dwelling in its infancy, this line of research aims to explore means for improving the accuracy of scoring positive and negative emotional tone of dream reports, before other dimensions can be explored.

In an early study, Nadeau, Sabourin, De Koninck, Matwin, & Turney (2006) used the Linguistic Inquiry Word Count software to identify and count occurrences of positive and negative words in dream reports as predictors of emotional positive and negative emotional tone. Using these frequencies along with their mathematical transforms (sqr, sqrt and log), they were able to produce the first model that could reliably predict the human judge's score of the dream report's negative emotional tone on a 1 to 4 scale. They had developed a model with an accuracy of 48%; nearly two times better than chance. Since then, Razavi et al. (2008), Amini, Sabourin & De Koninck (2009; 2010) and Matwin, Razavi, De Koninck & Amini (2010) have respectively introduced text-mining, dreamers emotional ratings, emotional progressions, word-associations, adverb-modified word vectors and co-occurrence vectors with advance machine learning techniques. These advancements have improved the accuracy of the machine model by extracting and utilizing novel and relevant sets of data from the text source. The latest model's accuracy in human-machine agreement was 67.3% (kappa 0.55) on the negative scale, and

78.7% (kappa 0.54) on the positive affect scale. This improvement is attributed to expansion of the machine learning and data extraction aspects of the Nadeau et al. (2006) paper.

Another aspect of the original research by Nadeau et al. (2006) that was not incorporated in the latest machine learning model was the data normalization through mathematical transform of attributes. With the progress in the machine learning aspect of this endeavor, the volume of extracted data has also expanded, and is increasingly difficult to manage. Tracking of the volumes of data is one difficulty that faces memory restrictions and data organization. Also, processing of these additional data also requires processing power and time. These have been the main reasons why the mathematical transforms of data have not been included. With advancement in computer memory and processing technology, these obstacles can be overcome. In the current study, this obstacle was challenged by reproducing the latest model with the addition of the mathematical transforms of the attributes. The latest model was based on text-mining, dreamer's emotional ratings, emotional progressions, word-associations and adverb-modified word vectors. It was predicted that the inclusion of the transformations will result in a more accurate model.

In addition to overcoming these computing limitations of the past, this research has sought to make the future of this research solely dependent on the dream report. In the history of improving the machine learning process, Razavi et al. (2008) utilized the dreamer's own rating of emotional tone as a source of data. Although this initiative was successful, it made the automatic analysis of dreams dependent on the dreamer's own rating on emotions; namely apprehension, anger, sadness, confusion, happiness, joy, fear and anxiety. This dependence imposes a requirement on the dream collection process that may make it difficult to utilize the model across different laboratories. To explore a purely textual analysis, it was decided to conduct the same experiment without the dreamer's own ratings of emotions. Without the predictive power of the dreamer's own ratings in the automatic scoring of positive and negative affect, it was expected that there would be more room to improve and thus the improvement should be more pronounced. Nevertheless, it was expected to demonstrate an improvement in performance due to data normalization through the mathematical transformations utilized by Nadeau et al. (2006).

Methods

The Canadian normative study of dreams conducted at University of Ottawa is an on going study that collects participant's account of their dream experience upon awakening at their homes. Participants are also asked to respond on a number of questions related to each dream; one of which is the dreamer's own rating of emotions. From this study 458 English dream reports met the 50-800 word count criteria and had a complete set of the dreamer's own rating of emotions. These dream reports were also scored on a 1 to 4 positive and negative affect scale by an experienced and trained human judge (author C.S). These scores are used to train and test the machine model. These dream reports and accompanying human judge scores were used in this study.

Prior to application of the machine learning techniques, dream reports are processed and attributes are extracted which fuel the learning process. Text-mining, dreamer's emotional ratings, emotional progressions, word-associations, adverb-modified word vectors were sources for extraction of attributes from the dream reports. Accompanying this set of attributes, the dreamer's own rating of dream's emotional tone (apprehension, anger, sadness, confusion, happiness, joy, fear and anxiety), were obtained from the dream report questionnaires. These attributes were used to build a Simple Logistic Regression model under the machine learning paradigm and represented the latest model. The machine learning process utilized 66% of the dream reports for training. In this stage appropriate attributes were selected and used to build a logistic regression using the human judge scores on positive and negative affect. The remaining 34% of reports (hold out set) were used to test the accuracy of the model against he human-judge's scores. The performance of the model was evaluated with a 10 fold cross-validation to ensure that random selection of dream reports did not bias performance.

A second model was built using the same process which included the square, square-root and logarithmic transforms of all the data. The performances of these two models were compared to quantify the contribution of the transformed data.

Finally, to test the performance of data transformations in the absence of dreamers' emotional ratings, the previous two steps were repeated with the exclusion of the dreamer's own emotional ratings.

Results

For the positive scale, when the dreamer's ratings were included, the performance peaked at 78.7% (kappa=0.54) and the inclusion of the transformed data did not contribute. With the dreamer's ratings removed, the performance improvement was 3.9%; from 71.6% (kappa=0.34) to 75.5% (kappa=0.43). For the negative scale, when the dreamer's ratings were included, the performance improved 1.9%; from 67.3% (kappa=0.55) to 69.2% (kappa=0.57). When the dreamer's ratings were excluded there was a performance improvement of 0.7%; from 61.5% (kappa=0.46) to 62.2% (kappa=0.47).

Discussion

We see that inclusion of the transformed data certainly does improve the performance of the model as expected. For the positive scale, there was an improvement in performance only when the dreamer’s own ratings were removed and the pronounced effect of transformed data was evident in the absence of the dreamer’s own ratings as hypothesized. This implies that data transformations seem to account for some of the predictive power of the dreamer's own ratings of emotions and perhaps one day, the dreamer's own ratings may be predicted with better extracted attributes and their transforms. On the negative scale, we see improvement in both models; with or without the dreamer's ratings as he had expected. The pronounced effect was not evident in the absence of dreamer’s emotional ratings. All of these findings indicate that a model may be employed solely on the dream reports which can allow this model to be used across different research laboratories that do not have a record of the dreamer ratings. Finally, it should also be noted that positive affect scores from the human judge were skewed compared to the negative affect scores. In the future, once a scale of a finer gradient is implemented, those scores may be normalized to better the machine’s performance.

References

Amini, R., Ouellet, D., Sabourin, C., & De Koninck, J. (2009). Word Associations Improve Automatic Analysis of Dream Emotional Tone. *Vigilance,* 19, P038.

Matwin, S., Razavi, A., De Koninck, J., & Amini, R. (2010). Classification of Dreams using Machine Learning. *Proceedings of the ECAI 2010*, 169-174.

Nadeau, D., Sabourin, C., De Koninck, J., Matwin, S., & Turney, P.D. (2006, July). *Automatic dream sentiment analysis*. Poster session presented at the Twenty-First National Conference on Artificial Intelligence (AAAI-06), Boston, Massachusetts

Razavi, A. H., Amini, R., Sabourin, C., Sayyad Shirabad, J., Nadeau, D., Matwin, S., & De Koninck, J. (2008). Evaluation and Time Course Representation of the Emotional Tone of Dreams Using Machine Learning and Automatic Text Analyses. *Journal of Sleep Research*, *17*, 257.

2.3 Article 3

AUTOMATIC DETECTION OF GENDER IN DREAM REPORTS

Christina Wong, Reza Amini & Joseph De Koninck PhD

School of Psychology, University of Ottawa, Ottawa, Ontario, Canada.

Abstract

A computer program called Automatic Analysis (AA) was developed to differentiate the dreams of males from females. Hypothesized gender predictors were based on previous literature concerning both dream content and written language features. The 144 male and 144 female dreams from 100 male and 100 female adolescent Anglophones were matched for equal length. Two male and two female undergraduate students were additionally asked to read all 288 dreams and guess the dreamer's gender. Results of the human judges indicated a pairwise percent correct gender prediction ranging from 70.6% (κ= 0.41) to 77.9% (κ= 0.557), for an average of 74.6%. Results from AA indicated a percentage of gender predictability of 74.3%, κ= 0.492 (chance is 50%, with κ= 0.00), demonstrating that AA can successfully predict the gender of written dream reports.

Introduction

Modern scientific study of dreams relies mainly on quantitative and objective approaches to dream content. Quantitative content analysis organizes recurring dream elements into specific categories from which frequencies are turned into percentages and rates. The scores of individuals are compared with norms in order to study the patterns in dream content of specific individuals and groups. One of the most elaborate and well-established methods to analyze dreams with is the coding system developed by Hall and Van de Castle (1966). In brief, this scale has unique codes for different aspects of dream content, as well as for the dreamer. For example, there are codes for different categories of dream content like characters (further subdivided into humans, animals, and imaginary figures/creatures), emotions (happiness, sadness, anger, apprehension, and confusion), activities (e.g., physical versus non-physical), social interactions (friendly, aggressive, and sexual), physical surroundings (settings and objects), strivings (failure and success), and misfortunes/good fortunes. A detailed set of up-to-date Hall & Van de Castle scale (HVdC) coding rules, tools, and examples is readily available online, along with an efficient spreadsheet used for statistical analysis called DreamSAT (Schneider & Domhoff, 2009). Study results may also be compared to normative findings (Hall & Van de Castle, 1966). In dream analysis techniques such as this one, human judges are responsible for coding individual dreams before entering the data into a computer spreadsheet such as DreamSAT for analysis. Furthermore, as an overall measure of objectivity, inter-rater reliability scores are calculated using the scores of multiple judges.

As such, the present study aims to analyze dream content with an automatic technique (Amini, Sabourin, & De Koninck, 2009) in order to predict the dreamers' gender. This technique, Automatic Analysis (AA), specifically refers to the combined processes of Linguistic Inquiry and Word Count (LIWC), word associations, progressions of LIWC categories and emotions, and the transformations of the data resulting from these processes (Amini, Sabourin, & De Koninck, 2011).

For this study in particular, the purpose of AA was to analyze written dream reports and produce numeric data, which were used by machine learning algorithms in a gender predicting learning task. In the initial stage, 66.7% of dreams were used to learn the gender classification task, while the remaining dreams were used to test the accuracy of the learned model. To resolve any bias that may have been introduced by the random sampling of dreams, a 10-fold cross-validation was applied. At each cross-validation, different random samples of dreams were used. Additionally, the reported percentage of agreement and kappa score designated the average performance of AA. For more details on machine learning, one may refer to the article by Matwin, Razavi, De Koninck, & Amini (2010).

To date, automatic techniques have been effectively used to determine the author gender of fiction and non-fiction book excerpts (Koppel, Argamon, & Shimoni, 2002), and to evaluate the overall emotional tone of dreams (Nadeau, Sabourin, De Koninck, Matwin, & Turney, 2006). When analyzing a dream's overall emotional tone, researchers generally look at particular expressions of positivity or negativity. Also, recent studies tend to focus more on the negative scale of dream emotion because of previous results that have shown that negative emotions occur more frequently in dreams and are also easier to differentiate (Roussy, Raymond & De Koninck, 2000). In their study, Matwin et al. (2010) showed that negative affect could be predicted at an accuracy level of 64%, which was comparable to the average agreement of human judges at 69%. Moreover, as opposed to solely searching through text for specific target words, more recent research by Amini, Sabourin, & De Koninck (in press) has been shown to improve the accuracy of automatic scoring with the incorporation of word associations. For example, "sadness" may also be associated with words like "crying", "tears", or "hurt", and such inclusions have been shown to improve the automatic scoring of the emotional tone of dreams.

As a next step in the development of the AA technique, it was felt relevant to attempt to apply the model on a longstanding issue: gender differences in dreams (Domhoff, 2005; Schredl, 2007). Hall and Van de Castle (1966) outlined a list of distinct gender differences in dream content that they believed were intrinsic, according to the dreamer's biological sex. For example, female dreams contain more characters and friendly interactions, while male dreams contain more aggression and achievement. These differences were observed in other studies (Winget, Kramer &Whitman, 1972) and found to be stable over time (Hall et al., 1982). However, more recent studies suggest that changes in social roles of women tend to reduce gender discrepancies in dreams (Lortie-Lussier, Schwab, & De Koninck, 1985; Lortie-Lussier, Simond, Rinfret, & De Koninck, 1992; Ouellet, Pérusse, Paquet-Biron, Sabourin, & De Koninck, 2009). Intercultural studies have also shown some evidence that gender differences may be dependent on the dreamer's culture. One such example includes the study by Urbina and Grey (1975) in which the dreams of female Peruvians were found to contain slightly more male characters than the dreams of male Peruvians, which contrasts with North American findings regarding the gender ratios of dream characters. Further yet, the significance of this dispute is also related to the heavily debated idea that dreams are reflections of the dreamer's waking life concerns and experiences (e.g., the influences of culture and society), which is in accordance with the continuity hypothesis of dreaming. Therefore, it was reasoned that the detection of gender differences using AA would not only contribute to the development of this technique but may also contribute to the broadening of our understanding of these gender differences.

To direct the detection of dreamers' gender, it was first necessary to quantify some of the relevant features that have been noted in previous research. In particular, both linguistic features and dream content aspects were investigated. The linguistic features were extracted by AA with the use of predefined LIWC categories, while the dream content features were quantified using the HVdC scale. The specific features, along with an elaboration of how they were quantified, are described below.

In summary, the main objectives of this study were to (1) quantify dream content for gender identification, (2) establish a baseline level of human accuracy in scoring for gender identification, and (3) utilize AA and apply a machine learning scheme, using data from existing scales to predict gender. Accordingly, it was hypothesized that (1) a human judge could predict a dreamer's gender from dream reports with good accuracy, (2) AA would predict a dreamer's gender with an accuracy level comparable to a human judge using dream content and written language features, (3) machine learning could effectively predict dreamer gender using relevant elements of the HVdC scale, and (4) a combined set of aspects, as extracted by AA and HVdC, could produce an even more accurate model.

Methods

Participants

Participants were drawn from a larger population of data collected in an on-going normative study of Canadian dreams at the University of Ottawa Sleep Laboratory. This study was initiated in 2005 and contains dream reports and surveys from Canadian males and females, both Anglophones and Francophones, from the ages of 12 to 65 years old and up. Recruitment of participants was done via word-of-mouth and poster advertisements on campus. All participation was voluntary and unpaid. Consent to participate was provided upon the signing of a consent form.

For the current study, the sample consisted of 100 male and 100 female adolescent Anglophone dreamers, ages 18 to 24 years old. This sample contained only the dream reports of Anglophones because the text-mining computer program in which the dreams were analyzed is currently only based on the English language (further details on this program are explained in the

Procedure section). To account for the general differences in word length among the dream reports of males and females (Hall & Van de Castle, 1966), only dream reports of a length within the range of 60 to 500 words were included in the study. More specifically, all of the male and female dreams were matched for dream length. As such, either one or two dreams were used per dreamer, for a total of 288 dream reports.

Additionally, four human judges (two males and two females, all within the age range of 18 to 24 years old) of the University of Ottawa undergraduate student population were asked to read all 288 dreams and to predict the dreamer's gender to their best of abilities. For their time, these judges were each compensated an amount of $50.

Materials

All participants were given a questionnaire battery in which they reported information concerning their first two dreams within a period of 10 consecutive days. From the questionnaire package, only the general information questionnaires and dream journals of Anglophones were used. The purpose of the general information questionnaire was to collect information concerning the participants' demographics (i.e., age, maternal language, sex, marital status, education, employment, and medications), sleep habits (usual sleep schedule, subjective insomnia, and daytime naps), and dream recall (frequency of dreams and nightmares). Upon awakening, participants used the dream journal to record their remembered dreams using as many details as possible to describe the locations, events, characters, interactions, activities, feelings, and emotions involved in the dream.

Measures

Linguistic Inquiry and Word Count (LIWC) is a software program designed by Pennebaker, Booth, and Francis (2001) to analyze various forms of texts (such as dream reports, book excerpts, emails, and transcribed speech). With the use of 64 categories of word features, the LIWC program can assess the degree to which people use different words in written language (for example, verbs, self-references, positive or negative emotions, and swear words). For the purposes of the present study, the English LIWC dictionary was used by AA to extract different features of written language and dream content. Additionally, AA utilized the target words and associated categories provided by this dictionary to search for associated words. Further details about the LIWC program and dictionary can be located online, along with links to the original English dictionary and its various translations (Pennebaker, Booth, & Francis, 2007).

Procedure

Upon receipt of the questionnaire package at the University of Ottawa Sleep Lab, each participant was given an alphanumeric identification code in order to maintain confidentiality. The dream summaries were then input into the computer with previous dream reports. From the collection of Canadian dreamers in the on-going normative study, 200 adolescent dreamers with dreams within the eligible word count limit (of 60 to 500 words) were selected. Within the total of 288 selected dreams, 144 adolescent female dream reports were matched for word count to 144 adolescent male reports. Two trained judges both then coded the dream content of each dream report according to the guidelines of the HVdC coding system (Domhoff, 1996). This same set of dream reports was used in *Parts A* to *D*.

Part A – Human Judge Scoring. The set of dream reports was randomly ordered in a document without any labels indicative of the dreamers' gender. A printed dream package was provided to each of the four human judges for the task of predicting the dreamer's gender. All judges were untrained in the areas of sleep, dreams, and gender differences, and were asked to identify dreamer gender based on dream report content. The judges' predictions of dreamer's gender were compared to the actual gender of the dreamer. Statistical measures of accuracy (percentage of agreement and kappa score) were calculated.

Part B – Automatic Analysis Scoring. Using the machine-learning scheme, 66.7% of the corpus of dream reports was used to learn and create a model, and the remaining dreams were used to test the performance of the model (AA.model). The final accuracy scores for gender prediction were determined after the 10-fold cross-validation process. In this process, Simple Logistic Regression was used as a means for classification (Matwin et al., 2010). Measures of accuracy included percentage of agreement between the actual gender and a kappa score. Cohen's kappa coefficient score is a measure of inter-rater agreement beyond chance. In this case, kappa represents the agreement between the predicted machine score and the dreamer's gender. These results regarding accuracy and reliability were then compared to those of the untrained human judges in order to assess the program's accuracy of gender prediction. Additionally, the significant features were compared to the traits that were hypothesized from literature to differentiate the genders.

In quantifying specific features that were suspected to be indicative of gender, we applied the infrastructure of the AA that was used to detect the emotional tone of dreams (Amini et al., in press). This model included some features that research has also recognized as being predictive of the dreamers' gender. Furthermore, Mulac, Bradac, and Gibbons (2001) asserted that written language features indicative of male dream reports include the increased use of first-person references and mentions of quantity. In the AA model, LIWC categories 4 (first-person references), 21 (numbers), and 20 (quantity), correspond directly to the language features that Mulac et al. (2001) identified. The total number of references to each category and the progression of these references were included in the analysis. Meanwhile, Mulac et al. (2001) also found that female dream reports appeared to incorporate more elaborate descriptions; as a measure of this feature, AA was programmed to include a sentence and word count for each dream report. In addition to these features, frequency of references to the remaining LIWC categories, along with their progressions, were included to enrich the dataset. Table 1 lists all 64 LIWC categories. Words and word associations were also included in this model as a measure of dream content.

Table 1. LIWC 2007 categories grouped and listed.

Total Function Words (Linguistic Processes)	**Psychological Processes**
Total pronouns	*Social processes*
Personal pronouns	Family, Friends, Humans
1^{st} person singular*, 1^{st} person plural, 2^{nd} person, 3^{rd} person singular, 3^{rd} person plural	*Cognitive processes*
Impersonal pronouns	Insight, Causation, Discrepancy, Tentative, Certainty, Inhibition, Inclusive, Exclusive
Articles	*Affective processes*
Auxiliary verbs	Positive emotion, Negative emotion, Anxiety, Anger, Sadness
Past tense, Present tense, Future tense	*Perceptual processes*
Adverbs	See, Hear, Feel
Prepositions, Conjunctions	*Biological processes*
Negations	Body, Health, Sexual, Ingestion
Swear words	*Relativity*
Common verbs	Motion, Space, Time
Quantifiers*	**Personal Concerns**
Numbers*	Work, Achievement, Leisure, Home, Money, Religion, Death
Spoken Categories	
Assent, Non-fluencies, Fillers	

Note. Further details can be found online at http://www.liwc.net/descriptiontable1.php. * = Categories that were expected to be predictive of gender, according to past research.

Part C – Hall & Van de Castle Scale Scoring. The same set of dreams was then manually coded by two trained human judges according to the HVdC scale for dream content. The judges consulted and came to a consensus on scores that were then entered into DreamSAT. Once in DreamSAT, the codes were converted into frequencies that were used to automatically predict the dreamer's gender under the machine learning paradigm (as previously described). Table 2 outlines all of the categories that were quantified using DreamSAT. In addition to these features, the frequencies of a few specific items from the objects category (i.e., tools and weapons, clothing, and household items) were manually counted because they were not automatically calculated in

DreamSAT. Furthermore, social aggression was calculated as total incidences of aggression minus physical aggression. From this set of data, only a subset of the features (as directed by literature) was used in the first model (HVdC.model1).

Table 2. DreamSAT counts of selected categories of the Hall and Van de Castle scale.

CHARACTERS*	**ACHIEVEMENT OUTCOMES**
Male*, Female, Unfamiliar, Familiar, Human, Animal, Friend, Family, Group	**Success**
	Dreamer involved
	Failure
SOCIAL INTERACTIONS	Dreamer involved
Aggression*	**ENVIRONMENTAL PRESSURE**
Dreamer-involved aggression*, Witnessed aggression, Dreamer as aggressor*, Dreamer as victim, Reciprocal aggressions, Mutual aggression, Self-directed aggressions, Male involved, Female involved, Physical aggression*	**Misfortune**
	Dreamer involved, Physical injury/Physical obstacle
	Good Fortune
	Dreamer involved
	EMOTIONS*
	Negative, Dreamer involved, Dreamer involved & negative
Friendliness	**SETTINGS**
Dreamer-involved friendliness, Witnessed friendliness, Befriender, Befriended, Reciprocal friendliness, Mutual friendliness, Self-directed friendliness, Male involved, Female involved	Indoor*, Outdoor*, Familiar*, Unfamiliar*
	OBJECTS*
	Body parts, Specific body parts (torso, anatomy, sex organs)
Sexuality	**DEAD**
Dreamer-involved sex, Witnessed sex, Dreamer as initiator, Dreamer as recipient, Reciprocal sex, Mutual sex, Self-directed friendliness, Male involved, Female involved	Dead characters
	STRIVING
	Success + Failure
	OVERALL POSITIVITY OF DREAM
ACTIVITIES	Friendliness, Success, Good Fortune
Dreamer involved, Physical*, Dreamer-involved & physical, Movement, Location change, Visual, Auditory, Verbal, Expressive communication, Thinking	**OVERALL NEGATIVITY OF DREAM**
	Aggression, Failure, Misfortune

Note. * = Categories that were expected to be predictive of gender, according to past research.

With regards to dream content, the literature suggested that male dream reports would include more cases of overall and physical aggression (Domhoff, 2005; Hall & Van de Castle, 1966; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003; Schredl, Sahin, & Schäfer, 1998); more incidences as the aggressor (Hall & Van de Castle, 1966; Krippner & Weinhold, 2002; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003); more mentions of male characters than female characters (Domhoff, 2005; Hall & Van de Castle, 1966); more references to outdoor (Domhoff, 2005; Hall & Van de Castle, 1966) and unfamiliar settings (Domhoff, 2005; Hall & Van de Castle, 1966); and more mentions of tools and weapons (Domhoff, 2005; Hall & Van de Castle, 1966; Krippner & Weinhold, 2002; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003). Meanwhile, female dreams were expected to include more incidences of social aggression (but less overall aggression) (Domhoff, 2005; Hall & Van de Castle, 1966); more mentions of characters (overall (Hall & Van de Castle, 1966) and familiar characters (Domhoff, 2005; Hall & Van de Castle, 1966; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003)); more references to indoor (Domhoff, 2005; Hall & Van de Castle, 1966; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003; Schredl, Ciric, Bishop, Gölitz, & Buschtöns, 2003) and familiar settings (Domhoff, 2005; Hall & Van de Castle, 1966); more descriptions of clothing and household objects (Domhoff, 2005; Hall & Van de Castle, 1966; Krippner & Weinhold, 2002); and more mentions of emotions (Domhoff, 2005; Hall & Van de Castle, 1966). In Table 2, this selection of gender-differentiating features is highlighted amidst a summary of all quantified and existing features in the HVdC scale.

To explore a more comprehensive model, a second model (HVdC.model2) was built using all of the HVdC scale features. This model was subjected to identical feature selection and testing as HVdC.model1. The percentage of accuracy and kappa scores of these two models were then compared to the results achieved for the AA.model.

Part D – Hall & Van de Castle Scale and Automatic Analysis Combined. Since the AA model was founded on linguistic features, while the HVdC models were founded on coded dream content, it was deemed important to investigate if a combined model would have a cumulative effect. In this endeavour, the significant features of AA.model were combined with those of HVdC.model2, and Combined.model resulted. Combined.model was tested under the same machine learning paradigm as the previous models. The percentage of accuracy and kappa scores of this model were calculated and compared to that of the untrained human judges.

Results

Part A – Human Judge Scoring

The results of the human judge scoring are listed in Table 3. Their accuracy ranged from 70.6% to 77.9%; with an overall average of 74.6% correct gender prediction. The corresponding Cohen's kappa (κ) measures of diagnostic reliability ranged from 0.41 to 0.557 and were within the range of moderate level of agreement. The results of all four judges were greater than the chance probability of agreement of 50% (κ= 0.00). Therefore, these results permitted the establishment of a baseline human accuracy in the scoring of gender identification. The machine's level of accuracy in gender prediction was compared to this baseline.

Table 3. Results regarding the accuracy and reliability of human judge scoring for correct dreamer gender prediction.

Human Judges	Accuracy	Kappa	Level of Agreement
Judge 1	70.6%	0.41	.41 to .60 as moderate
Judge 2	74.7%	0.492	.41 to .60 as moderate
Judge 3	75.1%	0.502	.41 to .60 as moderate
Judge 4	77.9%	0.557	.41 to .60 as moderate

Note. Chance agreement is 50%, kappa=0.

Part B – Automatic Analysis Scoring.

Table 4 summarizes the results achieved using AA. The results of AA (with the LIWC, word associations, progressions, as well as the logarithmic, square, and square roots of each of these three attribute classes) demonstrated an accuracy of 74.3% in successful gender identification with a Cohen's kappa of 0.492, indicating moderate agreement. Of the attributes extracted by AA, the gender-distinguishing traits that were retained in the final model were considered to be significant gender predictors. More specifically, the LIWC categories that were found to be significant gender predictors were linguistic processes (with increased usage of third-person pronouns, discrepancy, prepositions, and quantifiers), psychological processes (words related to cognitive mechanisms and friends), and personal concerns (words relating to leisure, home, and achievement).

Table 4. Results regarding the accuracy and reliability of multiple machine learning analyses for correct dreamer gender prediction.

Source of Data	Accuracy	Kappa	Level of Agreement
AA	74.3%	0.492	.41 to .60 as moderate
HVdC (all elements)	62.4%	0.263	.21 to .40 as fair
HVdC (selected elements)	60.2%	0.205	.00 to .20 slight agreement
AA + HVdC (all elements)	74.3%	0.492	.41 to .60 as moderate

Note. Chance agreement is 50%, kappa=0.

Part C – Hall & Van de Castle Scale Scoring.

The results of the HVdC scoring were also analyzed with machine learning and are similarly outlined in Table 4. In the first model (HVdC.model1), a selected set of HVdC scale elements were included based upon suggestions in the literature, and the resulting accuracy was 60.2% (κ= 0.205, slight agreement). From this model, two items remained in the model as significant gender predictors – number of female characters and ratio of male to female characters. More specifically, the results indicated that female dreamers tend to have more female characters than male dreamers, along with a higher male-to-female character ratio.

For the second model (HVdC.model2), an accuracy of 62.4% correct gender prediction was attained, with a Cohen's kappa (κ= 0.263) indicative of fair agreement. The traits that remained in the model and contributed to the identification of male dreamers included increased instances of misfortune (especially involving the dreamer) and negative outcomes (such as failure and having a character interfere with a dreamer's success). Meanwhile, the dreams of females tended to contain more mentions of male characters, animal characters, and verbal activity, and these particular traits remained in the final model. These findings regarding both male and female dreamers had not been previously anticipated.

Part D – Hall & Van de Castle Scale Scoring Combined with Automatic Analysis.

The machine learning was further applied to Combined.model (a combination of aspects from both AA and the HVdC scale) in order to evaluate if an incremental improvement in accuracy could be observed. The results from Combined.model mirrored the percent agreement and kappa scores of AA.model (74.3%, κ=0.492). In addition, this model was identical to the model achieved in *Part B*, even though it did not include any of the aspects that were coded using the HVdC scale.

Discussion

This study has demonstrated that dream content can be effectively used to predict the dreamer's gender at a level of accuracy greater than chance. According to the Cohen's kappa scores, human judges exhibited a moderate accuracy in predicting the dreamer's gender. The judges' average score was used as a baseline to compare the models that were subsequently investigated. More specifically, dream content was coded using the HVdC scale and quantified using AA (respectively and in combination), and these methods provided sufficient information for artificial intelligence learning algorithms to build models with an accuracy close to the baseline scores. The human judges performed slightly better than the models because they were able to assess attributes of a dream report that the machine models were not yet programmed to incorporate (e.g., usage of adjectives, adverbs, etc.), and judges were able to look at the overall tone or voice of the entire dream report. Despite this, the judges acknowledged the ambiguity of the majority of dream reports and admitted to first making their decisions based on traditional gender stereotypes (e.g., a dreamer's mention of having a boyfriend and increased descriptions of surroundings indicated a female dreamer, while the use of sentences that were more to-the-point and expressed less emotions were indicative of a male dreamer).

In this exploration to see if dream content could be predictive of the dreamers' gender, previous research findings were used to supplement the construction of the models. Of all the traits that were hypothesized to be gender-distinguishing, only one was observed in our final models as a contributor: males made more references to quantity. Additionally, some unanticipated attributes were revealed. For example, males reported more misfortunes, mentions of leisure activities, discrepancy, cognitive mechanisms, and achievement. Meanwhile, females reported more male characters and animal characters.

Failure to observe the expected traits as contributors may have been specific to the given sample of dream reports (adolescent age group), or may perhaps be due to a societal shift experienced since the past publications that were referenced. More specifically, the gradual decline of gender distinctions in waking life societal roles over the past few years may have influenced this corresponding change in dream content (Lortie-Lussier et al., 1992). Also, both age group and social roles have been known to impact dream content (Rinfret, Lortie-Lussier, & De Koninck, 1991), suggesting that the hypothesized gender differences may have been more salient in the age groups that were not investigated in this study. Further yet, the AA model for gender predictions presented in this study is only a resulting combination of the most significant gender predictors. In other words, some features may have successfully predicted gender in early models but did not show up in the final model, and this suggests that their ability to predict gender was less accurate than the features that remained in the final model. Studying the contributions of each of these variables was outside the scope of this study that merely aimed to illustrate that such differences are existent and measurable. It is therefore warranted to conduct more extensive and thorough research to examine each of these traits and their contributions to gender differentiation.

A further, more general limitation of dream studies is that they remain forever restricted by the subjectivity of individual dream experiences and by the limitations of the human language. Dream experiences are often very different from the experiences described in dream reports, especially as dreamers usually only report the events that take place with minimal details, while mostly neglecting to describe the intensity and greater details of the dream events (Strauch & Meier, 1996). Dream recall is also infamous for being imperfect (Moorcroft, 2005) and, as such, researchers are still only indirectly assessing dreams through written reports and by how the dreamer verbally communicates their visual dream experiences. Additionally, the HVdC coding system is still fairly subjective despite the outlined coding categories, as dream experiences are infinitely more complex and entail many situations that arise in dream reports that are not easily classified. Furthermore, in this study, it has been shown that the automatic scoring method of AA, free of reliance on human scoring, has outperformed all other models built on the HVdC scale. This suggests that the data extracted by AA may be encompassing or may be different than those coded by the HVdC as it pertains to gender-specific content. To clarify this matter, future research should use AA features to predict those of HVdC in order to see if AA features encompass the gender predicting features of HVdC.

Future studies may improve the accuracy of AA by quantifying additional attributes not yet included in the LIWC but that have been shown to distinguish genders (e.g., Domhoff, 2005; Hall & Van de Castle, 1966; Mulac et al., 2001). For instance, males appear to demonstrate an increased use of judgmental adjectives ("He seemed nice", "She was annoying") and elliptical sentences ("Nice sunset."); meanwhile, female gender predictors include an increased use of uncertainty words ("kinda", "maybe"), and intensive adverbs ("really", "very"), adjectives ("pretty", "golden") (Mulac et al., 2001). Automatic techniques may also be applied to the prediction of a dreamer's gender in other age groups, or to the prediction of a dreamer's age from a written dream report, as differences between age groups have been observed in the literature (e.g., Rinfret, Lortie-Lussier, & De Koninck, 1991). Future content analyses may consider analyzing data from women separately from men; the benefit of this approach would be to separate the variance in content due to gender from that which is of interest. Such gender-specific models may be useful for future dream analyses, especially in regards to emotional tone (such as in Amini et al., in press).

Presently, researchers in the area of machine learning are successfully combining research from different areas regarding language predictors of gender, age, native language, and personality, to automatically profile the authors of anonymous texts (Argamon, Koppel, Pennebaker, & Schler, 2009). Such large-scale machine learning findings have great implications for future studies with dream analysis, as this shows that machine learning models are indeed capable of combining multiple models of varying attributes and predictors into a larger model with an overall goal. The next steps with automatic techniques and dream analysis could potentially involve other aspects of dreamer demographics, dreamer psychopathology, written language features, and HVdC scale components, to be able to eventually mimic a standard computer model for the entire HVdC coding system. Such models may be combined from newly collected dream reports or from pre-existing online dream banks (e.g., Domhoff & Schneider, 2008) to construct a standard form or model from which anonymous dreamers may be profiled.

Ultimately, it seems that underlying gender distinctions in written language features and dream content still appear to persist despite the differing methods of analysis and extraction. As such, and especially in combination with the automatic techniques available for data analyses, the possibilities for future dream research have been expanded and may now be more easily adapted to areas beyond research in sleep and dreams.

References

Amini, R., Sabourin, C., & De Koninck, J. (2009). Word Associations Improve Automatic Analysis of Dream Emotional Tone. Poster presentation at The 4th Conference of the Canadian Sleep Society in Toronto.

Amini, R., Sabourin, C., & De Koninck, J. (2011). Transformed Attributes Facilitates Automatic Scoring of Emotional Tone of Dreams. Sleep, 34, A112.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2009). Automatically profiling the author of an anonymous text. Communications of the ACM, 52, 119-123. Domhoff, G. W. (1996). Finding Meaning In Dreams: A Quantitative Approach. New York, NY: Plenum Press.

Domhoff, G. W. (2005). The Dreams of Men and Women: Patterns of Gender Similarity and Difference. Retrieved February 4, 2011, from: http://dreamresearch.net/Library/domhoff_2005c.html

Domhoff, G. W., & Schneider, A. (2008). Studying dream content using the archive and search engine on DreamBank.net. Consciousness and Cognition, 17, 1238-1247.

Hall, C. S. & Van de Castle, R. L. (1966). The content analysis of dreams. New York: Appleton-Century-Crofts.

Krippner, S., & Weinhold, J. (2002). Gender Differences in a content analysis study in 608 dream reports from research participants in the United States. Social Behavior and Personality, 30, 399-410.

Koppel, M., Argamon, S. & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. Lit Linguist Computing, 17, 401-412.

Lortie-Lussier, M., Schwab, C., & De Koninck, J. (1985). Working mothers versus homeworkers: Do dreams really reflect changing roles of women? Sex Roles, 12, 1009-1021.

Lortie-Lussier, M., Simond, S., Rinfret, N., & De Koninck, J. (1992). Beyond sex differences: Family and Occupational Roles' Impact on Women's and Men's Dreams. Sex Roles, 26, 79-96.

Matwin, S., Razavi, A., De Koninck, J., & Amini, R. (2010). Classification of Dreams using Machine Learning. Proceedings of the ECAI-2010, 169-174.

Moorcroft, W. H. (2005). Understanding Sleep and Dreaming. New York, NY: Springer Science+Business Media. Mulac, A. (1998). The Gender-Linked Language Effect: Do Language Differences Really Make a Difference? In D. J. Canary & C. Dindia (Eds.), Sex Differences and Similarities in Communication: Critical Essays and Empirical Investigations of Sex and Gender Interactions (pp.127-153). Mahwah, NJ: Lawrence Erlbaum Associates.

Mulac, A. (1998). The gender-linked language effect: Do language differences really make a difference? in D. J. Canary and K. Dindia (eds), Sex Differences and Similarities in Communication: Critical Essays and Empirical Investigations of Sex and Gender in Interaction. Mahwah NJ: Lawrence Erlbaum Associates. pp. 127-53.

Mulac, A. (2006). The Gender-Linked Language Effect: Do Language Differences Really Make a Difference? In D. J. Canary & C. Dindia (Eds.), Sex Differences and Similarities in Communication: Critical Essays and Empirical Investigations of Sex and Gender Interactions (2nd ed.) (pp.219-239). Mahwah, NJ: Lawrence Erlbaum Associates.

Nadeau, D., Sabourin, C., De Koninck, J., Matwin, S., & Turney, P.D. (2006). Automatic dream sentiment analysis. Poster session presented at the Twenty-First National Conference on Artificial Intelligence (AAAI-06), Boston, MA.

Ouellet, D., Duchesne-Pérusse, A., Paquet-Biron, M., Sabourin, C., De Koninck, J. (2009). Disappearance of gender differences in the dreams of university students. Sleep, 32, A424.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count. Retrieved March 22, 2011, from http://liwc.net/

Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). Linguistic Inquiry and Word Count: LIWC2001. Mahwah, NJ: Erlbaum Publishers.

Rinfret, N., Lortie-Lussier, M., & De Koninck, J. (1991). The dreams of professional mothers and female students: An exploration of age and social roles impact. Dreaming, 1, 179-192.

Roussy, F., Raymond, I. & De Koninck, J. (2000). Affect in REM dreams: Exploration of a time-of-night effect. Sleep, 23, A174-A175.

Schredl, M. (2007).Gender Differences in Dreaming. In D. Barrett & P. McNamara (Eds.), The New Science of Dreaming (pp.29-47). Westport, CT: Praeger Publishers.

Schredl, M., Ciric, P., Bishop, A., Gölitz, E., & Buschtöns, D. (2003). Content analysis of German students' dreams: Comparison to American findings. Dreaming, 13, 237-244.

Schredl, M., Sahin, V., & Schäfer, G. (1998). Gender differences in dreams: Do they reflect gender differences in waking life? Personality and Individual Differences, 25, 433-442.

Schneider, A., & Domhoff, G. W. (2009). The Quantitative Study of Dreams. Retrieved February 4, 2011, from http://www.dreamresearch.net/

Urbina, S. P. & Grey, A. (1975). Cultural and sex differences in the sex distribution of dream characters. Journal of Cross-Cultural Psychology, 4, 374.

3. Summary of Results

The summary of results of the three articles is presented in this section.

Our the first article entitled "Word Associations Contribute to Machine Learning in Automatic Scoring of Degree of Emotional Tones in Dream Reports", attempted to show that word definitions from external sources could replace proximal co-occurrence vectors in efforts to improve the automatic scoring of emotional tone of dream reports. This article showed that an automatic system could reliably score for negative affect on a 4 level scale at an accuracy of 62.5% (kappa = .485). Similarly it showed that positive affect could also be scored accurately at 76.9% (kappa = .52). These findings supported the article's hypothesis.

In our second article entitled *"Data Normalization Assists Automatic Scoring of Emotional Tone of Dreams"* it was shown that the performance of the automatic scoring of emotional tone in dreams could be further improved by normalizing skewed data. It was shown that, with the inclusion of the transformed data, accuracy in scoring negative affect on a 4 level scale improved to 69.2% (kappa = .57). Similarly, the performance on the positive affect scale also improved in accuracy to 78.7% (kappa = .54). Both of these results support the hypothesis that data normalization was necessary in improving the performance of the artificial intelligence model.

In the last article entitled *"Automatic detection of gender in dream reports",* predicting dreamers' gender using dream content was explored. Results from this article showed that human judges could predict the dreamer's gender at a much better rate than chance. The selected human judges exhibited an accuracy that ranged from 70.6 to 77.9% (kappa = .41- .557). The average inter-judge agreement score was 74.6% (Fleiss' kappa =

.491). The results were used as a bench mark score. The automatic scoring model showed an accuracy of 74.3% (kappa = .492). The results in predicting the gender of the dreamer as done by human or machine were very similar. These scores confirm that there are differences in dream content between male and female and were sufficient to aid the automatic system in predicting the dreamer's gender at nearly the same accuracy as the average human judge.

4. Discussions

In a scientific approach to dream content analysis, research has come a long way. With comprehensive systems such as HVdC, researchers have had means to code specific aspects of dreams and thereby quantifying qualitative data that dream reports represent. These scales have been the bridge between qualitative-descriptives and the statistics which have allowed researchers to compare groups' means on given dimensions of dream reports. In other efforts to further and broaden the scope of dream content analysis, computers have been utilized to automatically code the content of dreams through analysis of their textual report. This initiative has relieved human judges of hours of coding; efforts which have been prone to reliability issues, human bias and error. Furthermore, the computational power of computers has opened an opportunity to analyze greater volumes of data.

This thesis, comprised of three articles addressing word associations, data normalization and gender differences in dream content, are contributions which show promise for future research in the field of dream content analysis. Each article contributed by demonstrating how the performance of such an automatic tool might be improved and how its scope may be broadened.

4.1 Word association

In the article entitled *"Word Associations Contribute to Machine Learning in Automatic Scoring of Degree of Emotional Tones in Dream Report"*, it was demonstrated that external sources can be used to establish word associations, which may be applied to dream content analysis. The benefits of such an approach, over proximal word co-occurrence vectors (PWCV) built on dream content, rested in removing a confound. PWCV (based on dream reports) were directly related to the dreams which were later evaluated. In addition to this, improvements in performance, provided support for the existence of implicit data that were communicated through explicit words. The extra information that was communicated might be labeled as the meaning a word may convey through its implied associations to other words. This study looked at general word associations as found in publicly available definitions, which were presumed to represent common knowledge. Aside from the common meaning of words, it was noted that words have specific meaning to the dreamer that may differ from its common sense. But, due to the scope of this research, such endeavour was not undertaken and is left to future research. Nevertheless, the method to describe words as vectors of association has been described in this article. One implication is that single word searches can be replaced by searching for words with similar associations. In this way, research can focus on searching for meaning (words with similar associations). Searches for words related in meaning can be furthered expanded to group similarly related words and labeled them as new symbols. Another implication is that dream reports no longer have to be treated as bags of words (sparse and seemingly unrelated sets of words), but rather as a series of words with large networks of associations. In this expanded form of representation, artificial intelligent techniques such as clustering can be utilized to group dream reports that share similarities in content.

This form of textual representation, a sequence of word associations, has helped to better represent dream reports. One may also look at the relationship between words in the grammatical structure of sentences (Amini & De Koninck, 2011a). Reflected in a more accurate scoring of positive and negative emotional tone of dream reports, this model may broaden its scope to score specific emotions. Alternatively, this approach may be used to score elements of the HVdC system as means for cross-validation.

4.2 Transformed data

The second article entitled *"Data Normalization Assists Automatic Scoring of Emotional Tone of Dreams"*, has demonstrated that data, as extracted from dream reports, can benefit from normalization (Amini & De Koninck, 2011b). This was evident in the improvements in accuracy of scoring for emotional tone of dream reports, due to normalization of data. This research described the method by which large volumes of data might automatically be normalized and lend themselves with greater ease to automatic scoring of emotional tone of dreams. This improvement in accuracy was substantial and invites the opportunity to remove dreamer's own rating of emotional tone in efforts to focus the automatic scoring system to be solely dream report dependent.

Management of the increased volume of data was a challenge, and might be a problem with future expansions. To meet the requirement for data normalization, and reduce the strain of resource usage, one should consider an automated yet selective data normalization routine. Statistical measures of skew can be used to qualify normalization.

In this context, normalization may entail an iterative process where a good mathematical function may be selected from a set of functions in an effort to normalize the data.

It should further be noted that the quickest and most simple learning algorithm, Simple Logistic Regression, was used. The intent was to focus the research on the domain of data extraction from dream reports rather than the technicalities associated with selection and tuning of artificial intelligence algorithms. In the future, one may explore other learning algorithms such as decision trees and Support Vector Machine (SVM). With these routines, although demanding in processing time and resources, one may expect to see improvements in accuracy primarily due to the sophistication of the algorithms.

4.3 Detection of Gender

In the last article entitled *"Automatic detection of gender in dream reports"*, an initiative was taken to see if the gender of the dreamer can be reliably predicted using the content of dream reports under the paradigm of automatic scoring of dream reports. The results of this study showed that there were differences in content that could be used to accurately predict dreamer's gender with accuracy similar to that of human judges.

To have confirming evidence that there are differences in dream content between genders might serve as a guide for future scoring of dream content. For example, in studying emotional tone of dream reports, males might exhibit different content than females in portrayal of the same level on a given affect scale. In this way, studying emotional valence of each gender separately would introduce the possibility for a more accurate scoring. A comprehensive approach to this might be to include gender specific word associations to better represent words and their affiliations. One method might be to ask males and females for word definitions and use them to build word associations.

In this correlation study, gender of the dreamers is not a direct predictor of content. It is likely that gender is the mediating factor for a number of variables such as social roles (Lortie-Lussier, Schwab, & De Koninck (1985); Lortie-Lussier et al. 1992) and biological factors (e.g., Domhoff, 2005; Lortie-Lussier, Simond, Rinfret, & De Koninck, 1992; Schredl, 2007). The source of these differences might be useful to further direct this line of research in helping us to extract specific data pertinent to those sources. For example, people of similar age groups might have been exposed to a unique upbringing that might affect their understanding and use of words. In this case, it would be interesting to see if automatic scoring conducted on each age group separately would benefit from a more accurate representation and thus more accurate scoring. Other approach might be taken for people of different cultural groups, birth order, personality traits, and health.

5. Conclusion

In a systematic approach to studying dreams, research has investigated automatic means of scoring dream reports. Recently, text mining algorithms have shown promise in scoring emotional valence of dream reports. Such a tool has shown to be beneficial in providing means for scoring volumes of dream reports in a short time in a consistent way. Yet further improvement in the accuracy and broadening of the scope of such a tool continues to be necessary. The presented series of studies advance the field of automatic scoring in improving accuracy in scoring and broadening the field. These studies demonstrated the potential for future research using automatic analysis. Research may yet continue to implement psychological theories, such as emotions and memory, to extract meaningful information. With such developments, the field of dream research may have access to more information that may be used to explore aspects such as abstract expressions and common themes. With a detailed yet broad scope, the future of dream research facilitated by such tools may allow us to look at aspects of dreams that previously seemed impractical.

References

Amini, R., Sabourin, C., & De Koninck, J. (2011a). Adverb Modifiers Improve Automatic Scoring of Emotional Tone of Dreams. Sleep, 34, A112.

Amini, R., Sabourin, C., & De Koninck, J. (2011b). Transformed Attributes Facilitates Automatic Scoring of Emotional Tone of Dreams. Sleep, 34, A112.

Barcaro, U., Cavallero, C., & Navona, C. (2005). A Method for Recognizing and Describing the Links Among Dream Sources. Dreaming, 15(4), 271-287.

Benca, R. M., Obermeyer, W. H., Shelton, S. E., Droster, J. & Kalin, N. H. (2000). Effects of amygdala lesions on sleep in rhesus monkeys. Brain Research, 879, 130-138.

Cavallero, C. & Foulkes, D. (1993). Dreams as Cognitions. New York: Harvester Wheatsheaf.

Dang-Vu, T. T., Schabus, M., Desseilles, M., Schwarts, S., & Maquet, P. (2007). Neuroimaging of REM sleep dreaming. In Barrett, D., & McNamara, P. (Eds.), The new science of dreaming vol 1 (pp. 95-114). London: Praeger.

Domhoff, G. W. (2003). A new resource for content analysis. (pp. 95-105). Washington, DC, US: American Psychological Association.

Domhoff, G. W. (2005). The Dreams of Men and Women: Patterns of Gender Similarity and Difference. Retrieved February 4, 2011, from: http://dreamresearch.net/Library/domhoff_2005c.html

Foulkes, D. (1985). Dreaming: A Cognitive-Psychological Analysis. Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Hall, C., & Van de Castle, R. (1966). The Content Analysis of Dreams. New York: Appleton-Century-Croft.

Hartmann, E. (2007). The nature and functions of dreaming. In D. Barrett & P. McNamara (Ed.), The new science of dreaming: Volume 3 (pp. 171-192). Westport, CT, US: Praeger Publishers.

Koppel, M., Argamon, S. & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. Lit Linguist Computing, 17, 401-412.

Lortie-Lussier, M., Simond, S., Rinfret, N., & De Koninck, J. (1992). Beyond sex differences: Family and Occupational Roles' Impact on Women's and Men's Dreams. Sex Roles, 26, 79-96.

Lortie-Lussier, M., Schwab, C., & De Koninck, J. (1985). Working mothers versus homeworkers: Do dreams really reflect changing roles of women? Sex Roles, 12, 1009-1021.

Mancia, M. (2005). The dream between neuroscience and psychoanalysis. Schweizer Archiv fur Neurologie und Psychiatrie. 156(8), 471-479.

Matwin, S., Razavi, A., De Koninck, J., & Amini, R. (2010). Classification of Dreams using Machine Learning. Proceedings of the ECAI 2010, 169-174.

Nadeau, D., Sabourin, C., De Koninck, J., Matwin, S., & Turney, P.D. (2006, July). Automatic dream sentiment analysis. Poster session presented at the Twenty-First National Conference on Artificial Intelligence (AAAI-06), Boston, Massachusetts.

Razavi, A. H., Amini, R., Sabourin, C., Sayyad Shirabad, J., Nadeau, D., Matwin, S., & De Koninck, J. (2008). Evaluation and Time Course Representation of the

Emotional Tone of Dreams Using Machine Learning and Automatic Text Analyses. Journal of Sleep Research, 17, 257.

Roussy, F., Raymond, I. & De Koninck, J. (2000). Affect in REM dreams: Exploration of a time-of-night effect. Sleep, 23, A174-A175.

Schredl, M. (2007).Gender Differences in Dreaming. In D. Barrett & P. McNamara (Eds.), The New Science of Dreaming (pp.29-47). Westport, CT: Praeger Publishers.

Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: An empirically-based computational model. Cognitive Science: A Multidisciplinary Journal, 31(4), 645-671.