TEXT MINING DESCRIPTIONS OF DREAMS: AESTHETIC AND CLINICAL EFFORTS

Renato Fabbri[1] - renato.fabbri@gmail.com
Fabiane M. Borges[2] - catadores@gmail.com
[1]University of São Paulo, Institute of Mathematical and Computer Sciences - São Carlos, SP, Brazil
[2]Federal University of Rio de Janeiro, School of Fine Arts - Rio de Janeiro, RJ, Brazil

Abstract. *Dreams are highly valued in both Freudian psychoanalysis and less conservative clinical traditions. Text mining enables the extraction of meaning from writings in powerful and unexpected ways. In this work, we report methods, uses and results obtained by mining descriptions of dreams. The texts were collected as part of a course in Schizoanalysis (Clinical Psychology) from dozens of participants. They were subsequently mined using various techniques for the achievement of poems and summaries, which were then used in clinical sessions by means of music and declamation. The results were found aesthetically appealing and effective to engage the audience. The expansion of the corpus, mining methods and strategies for using the derivatives for art and therapy are considered for future work.*

Keywords: *Text mining, Dreams, Schizoanalysis, Poetry, Art*

1. INTRODUCTION

Although dreams are described in texts that range from ancient sacred (Barker et al., 1993; Boas & Boas, 1974; Kopenawa, 2013) to psychiatric (Esposito et al., 1999), there is no consensus of what dreams are. We can exemplify the diverse theories with three simple cases (Domhoff, 2013):

- dreams are often regarded by the dreamers as accessing spiritual realms and other realities.

- Many scientists regard dreams as by-products of the sleeping process: arbitrary interpretations given by the conscious mind to noisy signals without substantial meaning.

- Freudian and Jungian psychoanalytic traditions understand dreams as symbolic constructs output by the unconscious mind.

We can, however, state some facts about dreams that assert them as attractive for therapy and for art. First, dreams are often very rich in impacting and symbolic images. Second, they are told

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

by the person who dreamed in a very attentive manner, as being very significant to the dreamer. In fact, most of us should be able to recall a number of situations where someone (perhaps ourselves) was describing a dream in a rapid, almost euphoric, succession of words. Dreams are so effective in yielding artistic materials that surrealism is an aesthetic explicitly inspired by dreams and symbolism is an example of artistic movement heavily influenced by dreams.

Text mining is data mining applied to textual data. There are many models for the text mining pipeline, but it can be summarized as: data collection and preparation, pattern recognition, evaluation of the output and reporting (Fabbri, 2017). This work addresses text mining of descriptions of dreams with aesthetic and therapeutic purposes.

Section 2 describes the corpus and methods. Section 3 is dedicated to presentation and discussion of results. Section 4 holds conclusions and further work considerations.

2. MATERIALS AND METHODS

2.1 Corpus

The description of dreams we used are all in Brazilian Portuguese, collected as part of clinical practices in the year of 2015. Participants described the dreams and sent them to the second author of this paper, who is a psychologist. Thereafter, another collection of dreams was gathered in the same way by the second author and collaborators, with the purpose of expanding the analysis and synthesis of texts performed with the previous corpus. It is a larger corpus, also in Brazilian Portuguese. Interestingly, both corpus contain descriptions of dreams by women only. Their scales are summarized in Table 1 in terms of numbers of characters, tokens, and paragraphs.

Table 1: Corpus files and elementary countings. The number of dreams is about the same as the number of paragraphs. The date column corresponds to the month when the last dream was collected.

File	Characters	Tokens	Sentences	Paragraphs	Date
corpora.txt	9456	1693	104	30	Mar/2015
corpora2.txt	71514	14691	435	156	Nov/2015

2.2 Analysis and derivation methods

The texts were analyzed to support the extraction of meaning from the dreams and for the creation of artistic texts. We strived to keep the methods very simple in order to avoid puzzling the involved parties. Three lists of tokens were considered:

- punctuations !"#$%&'()*+,-./:;=¡=¿?@[]\^_'{}—˜. Obtained through the command `string.punctuation` of Python's string (standard) library.

- Portuguese *stopwords*[1] obtained through NLTK (Bird, 2006) by the command `nltk.corpus.stopwords.words("portuguese")`.

---

[1]The exact definition and list of stopwords are not consensual. Anyway, one can regard them as words with lesser meaning and which are very frequent, such as conjunctions and prepositions.

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

- Tokens in the texts which were not punctuations nor stopwords. These were regarded as the most meaningful words in the corpus.

This selection of most meaningful words was used as the core material for the achievement of more interesting constructions for art and clinical psychology through filtering and ordering. Most significantly, the ordering could be based on the alphabet, the size of tokens in number of letters, or the count of incidences of the words, or any combination of these. Filtering could be performed by restricting the vowels, consonants (e.g. fricatives), word size, frequency, or collocations.

3. RESULTS AND DISCUSSION

The list of most meaningful words (described in the previous section) was filtered and ordered in many ways to yield diverse sequences of interest. After an inspection of the results, these criteria were selected to compose a final document:

- Ordering by: incidence (most frequent words first), alphabetic, size in characters, with and without repetitions. These were considered the most raw sequences and used subsequently to derive other sequences with such variations of ordering and repetition.

- Words with only one vowel (repeated any number of times).

- Only words with fricatives or plosives or some combination of them (e.g. plosives and 'm' and vowels 'a' and 'e').

- Words that start and end sentences.

- Collocations (pairs of words which are frequent together).

Such final document and other files are available online and exposed in Table 2. An example of the derived texts is in Table 3 with a translation from Portuguese to English. These texts were used for aesthetic appreciation and also in a schizoanalysis group session in 2014 (Casa Nuvem, Rio de Janeiro, RJ, Brazil). In the same course, the artist Giuliano Obici used the texts to feed his program "Voices Simulacrum" (Obici, 2015): a machinic chorus of robotic voices with computers connected in network. The project integrated devices in the computer (sound, video and network cards) exploring the network as a distributed audiovisual instrument (conceptualized as a "metamedia-instrument") to read the texts in a very performative manner.

The group was constituted by the participants who described the dreams and the report of the episode is somewhat impressive: the members had strong impressions, some of them cried and entered a quasi-shock state.

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

Table 2: Files related to the text mining of dreams. All files are found in a public git repository dedicated to the developments presented in this article (Fabbri & Borges, 2015).

File	Description
`scripts/todos.py`	Python script that makes the current analysis and renders the TXT and PDF files.
`corpora/corpora.txt`	The first collection of descriptions of dreams.
`corpora/corpora2.txt`	The second (and larger) collection of descriptions of dreams.
`mineracaoDosSonhos/PLNSonhos.odt`	A brief consideration of the text mining of dreams to which this article is dedicated.
`mineracaoDosSonhos/TUDO.pdf`	A thorough exposition of all the (selected) texts derived from the descriptions of dreams.

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

Table 3: Example of artistic text achieved from the descriptions of dreams. This text was obtained through picking only the first and last words of each sentence.

Portuguese (original)	English (translation)
Escorregava glandes	Slipping glands
Numa assustavam	At once, they scared
Eu suada	I sweated
As cavalos	The horses
Não acabou	It's not over
Barras mim	Bars me
Andei construtores	I walked builders
Pessoas)	People)
Sonhei formei	I dreamed I formed
Estava menino	It was boy
Depois boa	Then good
Esse meu	This mine
Sonhos descendo	Dreams coming down
O irmão	The brother
Meu punição	My punishment
Começa irmão	Begins brother
Meu ele	My him
Meu demonstração	My demonstration
Depois "	After "
Eu parede	I wall
Sinto dele	I'm fell him
A importência	The "importence"
O buraco	The hole
Acordei ofegante	Woke up breathless
Sensação NÃO	Feeling NO
Já rumo	I'm on my way
Estava perseguido	Was persecuted
Quando percebeu	When realized
A tempo	In time
Até porta	Up to door
O disso	The this
Eu sobreviver	I survive
Parecia ferramentas	Seemed like tools
Três mim	Three of me
Lavo piano	I wash piano
Havia tirano	There was a tyrant
Nele tudo	In him all
Jogaram fogo	They set fire
Pessoas presas	People trapped
Comecei ali	I started there
Mas destruísse	But destroy
Pessoas criança	Children people

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

4. CONCLUSIONS AND FUTURE WORK

We understand that the results are compelling for both art and clinical psychology. Only the first corpus was used, which is smaller and made easier the selection of the resulting texts. The methods applied are very simple, favoring the communication between the parties, and are promptly deepened and expanded into more complex processes. This work seems unique in the sense of using text mining of dreams for art and clinical psychology, which, in our opinion, benefits the appreciation of it as a multidisciplinary and scientific contribution in computer science, art and psychology.

In further efforts, we might use for the corpus:

- descriptions of dreams in the literature (e.g. from the mythology, traditional communities, etc);

- other languages;

- an expansion of current corpus;

- dreams from specific groups, e.g. again gender related or of a specific age span, professional or educational background, etc.

About the text mining methods, we might:

- use specific routines for classification (e.g. clusterization) of texts or their features;

- expand the methods of selection of words to better encompass meter (e.g. number of syllables);

- expand the methods of selection of words and phrases by their sonorities (e.g. by using sequences of vowels or consonants, mute consonants, paroxytones);

- use Wordnet (Fellbaum, 2010) in order to relate terms through semantic links (e.g. hypernymy, meronymy, synonymy).

The exploration of the results in therapeutic sessions and for the achievement of collections of artistic texts should be kept as the core purposes.

Acknowledgements

REFERENCES

Barker, K. L., Burdick, D., Burdick, D. W. (1993). The NIV study bible, new international version. Zondervan Bible Publishers.

Bird, S. (2006). NLTK: the natural language toolkit. In Proceedings of the COLING/ACL on Interactive presentation sessions (pp. 69-72). Association for Computational Linguistics.

Boas, O. V., & Boas, C. V. (1974). Xingu: the Indians, their myths. Souvenir Press.

Domhoff, G. W. (2013). Finding meaning in dreams: A quantitative approach. Springer Science & Business Media.

XX ENMC e VIII ECTM
16 a 19 de Outubro de 2017
Instituto Politécnico Universidade do Estado do Rio de Janeiro – Nova Friburgo - RJ

Esposito, K., Benitez, A., Barza, L., Mellman, T. (1999). Evaluation of dream content in combat-related PTSD. Journal of traumatic stress, 12(4), 681-687.

Fabbri, R. (2017). Topological stability and textual differentiation in human interaction networks: statistical analysis, visualization and linked data. PhD thesis, University of São Paulo. `https://github.com/ttm/thesis/raw/master/thesis-rfabbri.pdf`

Fabbri, R., Borges, F.M. (2015). Public git repository with the texts, scripts and corpora for the analysis of dreams. `https://github.com/ttm/sonhos`

Fellbaum, C. (2010). WordNet. In Theory and applications of ontology: computer applications (pp. 231-243). Springer Netherlands.

Kopenawa, D. (2013). The falling sky. Harvard University Press.

Obici, G. (2015). Vozes Simulacrum: a machinic choir with connected computers. From `http://www.giulianobici.com/site/vozessimulacrum.html`