# Unraveling reported dreams with text analytics

Iris Hendrickx <iris_at_i-hx_dot_nl>, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands; Centre for Language and Speech Technology, Radboud University, Nijmegen, The Netherlands
Louis Onrust <l_dot_onrust_at_let_dot_ru_dot_nl>, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
Florian Kunneman <f_dot_kunneman_at_let_dot_ru_dot_nl>, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
Ali Hürriyetoğlu <a_dot_hurriyetoglu_at_cbs_dot_nl>, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
Wessel Stoop <w_dot_stoop_at_let_dot_ru_dot_nl>, Centre for Language and Speech Technology, Radboud University, Nijmegen, The Netherlands
Antal van den Bosch <a_dot_vandenbosch_at_let_dot_ru_dot_nl>, Meertens Institute, Amsterdam, The Netherlands

## Abstract

We investigate what distinguishes reported dreams from other personal narratives. The *continuity hypothesis*, stemming from psychological dream analysis work, states that most dreams refer to a person's daily life and personal concerns, similar to other personal narratives such as diary entries. Differences between the two texts may reveal the linguistic markers of dream text, which could be the basis for new dream analysis work and for the automatic detection of dream descriptions. We used three text analytics methods: text classification, topic modeling, and text coherence analysis, and applied these methods to a balanced set of texts representing dreams, diary entries, and other personal stories. We observed that dream texts could be distinguished from other personal narratives nearly perfectly, mostly based on the presence of uncertainty markers and descriptions of scenes. Important markers for non-dream narratives are specific time expressions. Dream texts also exhibit a lower discourse coherence than other personal narratives.

# 1. Introduction

Dreams are a fascinating phenomenon that has been studied for millennia. In ancient Greek and Egyptian times dreams were seen as messages from the gods, and played an important role in religion. One of the earliest recorded dream analyses was written on clay tablets in Mesopotamia, 5000 years ago [Black and Green 1992]. This ancient epic tale of Gilgamesh includes several dream descriptions and interpretations. Nowadays, various research fields still study the meaning and purpose of dreams, such as psychiatry, psychology, neuroscience, and religious studies. Despite all these research efforts, a comprehensive explanation of the purpose of dreams is still lacking today.

1

Psychologists and social scientists have studied dream content with quantitative methods for decades, working with the hypothesis that dreams reveal psychological information about the dreamer. One currently dominant theory in this area is the *continuity hypothesis*, which assumes that the content of dreams reflects a person's daily life and personal concerns [Domhoff and Hall 1996]. Previous studies on dream descriptions, i.e. reported dreams written afterwards by the dreamer, have shown that around 75–80% of dream content relates to everyday settings, characters, and activities. The remaining dreams are related to uncommon or even bizarre topics. Some of the latter are shared by numerous people, such as dreaming about flying, teeth falling out, or being naked in public [Domhoff and Schneider 2008].

2

Dream descriptions are written reports of the memories of an experienced dream. Even though much progress is made in neuroscience, it is not possible yet to decode the dream content from a dreaming person's brain activity. The only possible way to gather dream contents is to study the reported recollection produced when the experiencer was awake

3

[Domhoff and Hall 1996]. For this reason, we study written reports of remembered dreams. As a textual genre, this type of written report bears similarities to other written recollections of personal experiences, both in cognitive and sensory qualities [Kahan and LaBerge 2011]. In this study we aim to investigate what linguistic features are specific to dream reports, contrary to reports on personal stories that actually happened, using tools for automatic text analysis. Computational approaches to automatically analyze the content of dream reports from a linguistic perspective are rare. In this work we want to pave the road for further detailed and knowledge-directed research by presenting a first account of a computational text analysis of dream reports.

We performed three different types of automatic text analysis to investigate what typical characteristics we can discover in dream reports. We hope that our automatic linguistic approach can demonstrate to dream analysis experts how well-studied techniques from the field of computational linguistics can be applied to offer insights into linguistic patterns hidden in large dream collections. These analyses go beyond the standard word frequency analysis that is common in corpus linguistics and that has already been applied to dream reports [Hall and Van de Castle 1966] [Domhoff 2003].

The largest available digitally curated collection of dream reports is the DreamBank [Domhoff and Schneider 2008] which contains over 22 thousand dream reports gathered in the last decades in various scientific studies. We use DreamBank as the base for our study; we also collected a contrasting data sample of true personal stories (from Reddit and Prosebox) to perform our experiments.

We apply the following three methods: *automatic text classification* to investigate what features are actually salient for predicting whether a written text is a dream report or not, *topic modeling* to discover the common themes in the dream collection, and *text coherence analysis* to measure whether there is a difference in coherence between dreams and personal stories. Each of these methods offers a different perspective on dream data. As we will argue, they do lead to overlapping findings that we discuss in the last section. This paper is structured as follows. We first discuss related work in automatic textual analysis of dream reports and previous work on comparisons between dreams and stories in Section 2. We present the data sets used in the experiments in detail in Section 3. Next, we present the three different studies we have done in Section 4, and we summarize and discuss our findings in Section 5.

## 2. Related work

Automatic textual analysis of dream reports is a relatively unexplored field. Semi-automatic experiments have been performed by Bulkeley and Domhoff [2009], who developed a systematic category list of word strings that can be used for automated queries and word-frequency counts. The categories in which the words are organized relate to the content of dreams, and are used to count mentions of emotions, characters, perception, movement, cognition, and culture. In a more recent follow-up study [Bulkeley 2014] this category list is updated and evaluated on four data sets present in the DreamBank corpus. The study shows that one can use this type of word analysis to detect the general topics in dream content in the same way and with a similar accuracy as the more time-consuming manual analysis. Furthermore, Bulkeley offers evidence that based on an individual dream collection, it is possible to make accurate estimations about a person's life, his concerns, activities, and interests, thereby confirming the continuity hypothesis.

Some work exists on automatic text classification with machine learning methods, where the task is to assign emotion labels to dreams. In [Razavi et al. 2014], follow-up work to [Matwin et al. 2010], the authors aim to label dreams on a four-point negative/positive sentiment scale. The authors represent dreams as bag-of-word vectors and include dynamic features to represent sentiment changes in the dream story. They run ten-fold cross validation experiments on a sample of 477 manually labeled dream reports and achieve up to 64% accuracy, close to the average human agreement of 69%.

In [Frantova and Bergler 2009] a more refined type of sentiment analysis is explored; they predict the fuzzy assignment of five emotion categories to dream descriptions, based on semi-automatically compiled emotion word dictionaries. Their method is evaluated against a sample from the DreamBank that is manually labeled with the emotion annotations from the Hall/Van de Castle encoding system [Hall and Van de Castle 1966]. The difficulty of using these DreamBank annotations is that this labeling has been done at the document level, which is also the level at which [Frantova and Bergler 2009] evaluated their approach, even though the annotations refer to specific phrases in the dream. The direct

link between the linguistic description and label is missing.

The dream reports in the DreamBank were written down either by the dreamers themselves or by researchers that interrogated the dreamers after awaking. The written dream reports resemble oral narratives as they were described in the seminal work of Labov and Waletzky [Labov and Waletzky 1967]. They can be considered as spontaneously told short stories that relate an experience. Labov and Waletzky discuss the different crucial elements and structure in a narrative. A narrative consists of five or six structural units. It often starts with an *abstract*, which is a brief summary of the story, followed by an *orientation* that sketches the setting, location and actors of the narrative. The main unit of the narrative is the *complication* entailing the main series of events. Additionally, an evaluation unit expresses the attitude of the author towards the story, signaling the purpose of the story. The narrative is concluded with a *resolution* to complete the story and the *coda* that places the told story in a perspective and connects it to the current situation.

[Kilroe 2000] compared dreams to stories from a narrative perspective based on the ideas of Jung [Jung and Shamdasani 2010], and argues that some dreams have a clear narrative structure while others are just a fragment or a snapshot. A more specific definition is given by Montagero [Montangero 2012] who argues that the characters in a narrative need to have intentional states and that a narrative must introduce an unexpected event. He concludes that dreams indeed can be classified as narratives under this definition.

We also consider most dream reports as narratives, given that, in most cases, they are *tellable* [Labov 2006]. Although humans dream every night, the recollection of a dream requires some motivation and effort. Furthermore, the dream content is outside the dreamers control which makes it unpredictable and thus reportable. Dreams do contain the minimal part of a narrative [Labov 2008]: that of having *temporal juncture* and a complicating action where two events take place in a certain temporal order.

Drawing on the argumentation of this previous work we posit that dreams are personal narratives, which narrows down our research question to: what distinguishes dream reports from other personal narratives such as true stories?

# 3. Data

As we are interested in automatically investigating textual properties, and studying what characteristics are typical for dream reports, we compare dream reports to other texts and narratives. We use dream reports from the DreamBank, and we place them in contrast with data representing personal narratives that actually happened, taken from the internet sources Reddit and Prosebox. In this section we introduce the three sources, and describe their properties.

## 3.1. DreamBank

We use the dream reports from collections as gathered in the DreamBank, a project to combine the results of several scientific studies and resources over the years in one online search interface [Domhoff and Schneider 2008]. These collections of dream series vary greatly in type, size, and intended purpose. Some series consist of a longitudinal collection of dream descriptions of a single person, such as the collection "Dorothea: 53 years of dreams" consisting of around 900 dreams. Other series represent a specific group of dreamers such as adult male and female blind dreamers [Hurovitz et al. 1999]. Some collections are in English, gathered in Australia, Canada or the US. One of the collections, collected in Switzerland, contains dream reports in German. The DreamBank is an ongoing project and collections are added regularly. We use a snapshot from the DreamBank retrieved in April 2015 containing 22 thousand dreams divided in 67 different collections.

For our experiments we performed the following selection steps on the DreamBank data, where we limited ourselves to English written collections. Since some of the DreamBank collections overlap, we removed the duplicates from our sample. We also removed a part of the description in the collection "College women from the late 1940" that contained answers to specific questions, and we only kept the dream description itself. We applied an automatic language identification step [Lui and Baldwin 2012][1] that filtered out a handful of other dreams (for example dream #0694 of the Barb Sanders collection [Domhoff 2006] where she dreams about a Spanish conversation). Next, the data was tokenized automatically,[2] leading to a sample of 21,598 dream descriptions containing a total of 4.3 million word

tokens. Dream descriptions contain an average of 56 words, with a population standard deviation of 38.5.

We noted that some collections in the DreamBank are much larger than others, and that dream descriptions of certain persons (e.g. Barb Sanders) are relatively prominent in the DreamBank content. We decided to create a sample of the DreamBank where we limit the amount of dream reports per individual dreamer to a random selection of at most one hundred dreams. This produced a sample of 6,998 dream descriptions, comprising 1.3 million tokens in total with an average of 65 words and a standard deviation of 43.7 per dream description, similar to the larger sample. We used both the large and the small sample in our experiments.

17

We show an example of a tokenized dream description with 97 tokens:

18

I was chosen to be interviewed by S , the college president , but it's unclear if my papers were approved in time , so I clutch my briefcase with my acceptance letter in it and try to find the building .
A woman student and Ellie help guide me to the building .
I find a sign saying " 504 , " the room .
I rush to the room , hoping , feeling late and uncertain .
I am there in the nick of time .
I am calm and handle it well .

## 3.2. Personal Stories

To discover what the typical linguistic attributes of dream reports are, we need a comparable set of contrasting reports that is as similar to dream reports as possible, both in structure and in content. Comparing dream reports to a collection of news paper articles or personal letters will lead to obvious findings such as: dreams do not report on political debates and the weather forecast, and will not end in 'yours sincerely'. This is not the type of differences that we are interested in. We therefore aimed to find a collection of personally written recollections of true daily life events. Recall that dreams are known to reflect daily life events and activities for at least 75–80% of the cases.

19

Comparable collections of personal stories recollecting true events, not just fantasies or fiction, are difficult to find when looking for existing curated corpus collections. For this reason we resorted to collections of web texts to build our own corpus.

20

The first part of the contrasting data consists of personal stories. The stories are crawled from Prosebox,[3] an online community to share journals and personal stories. Just prior to this research, OpenDiary, a community where users could post diary entries, was taken down. Many of these users moved to Prosebox, and especially older posts are mostly diary entries or journals.

21

We collected all public posts that were available at the end of March 2015. As a result, we crawled 130 thousand posts with over 67 million tokens. We applied the same filtering pipeline to the Prosebox posts as was applied to the dreams; that is, we applied a language filter where we only kept the posts which were identified as English; second, we tokenized the posts. Since the number of tokens is much larger, we downsampled the corpus to a similar number of tokens as the DreamBank samples, i.e. the large sample and the smaller limited sample, containing 4.3 million words and 1.3 million words respectively, with an average of 64 and 63 words and standard deviations 78.8 and 94.3, respectively. In other words, we kept the average document size virtually equal to that of the DreamBank samples; the Prosebox data does exhibit a larger variance in size. We show an excerpt of a Prosebox text here:

22

Just sitting
Life is good here .
I had a good day of just staying home yesterday .
I went grocery shopping this morning and Cap is at his auction .
He has called me a couple of times and he is having a great time .
He loves seeing his friends that he sits with .
Tomorrow should be another day of staying home .
Yay .
If I lived by myself I wouldn't go anywhere .
I love staying home .
I bought groceries today .

I bought strawberries and whipping cream for strawberry shortcakes .

The second part of the contrast data consists of Reddit posts. Reddit is a website where users can submit content of almost every kind. The site uses a community system, where each community is called a subreddit. We collected posts from a number of subreddits where the posts are texts about daily and personal experiences such as communities named *offmychest*, *diaries*, *relationships*, *shortscarystories*, *lifeinapost*, *anxiety*, and *self*. Prior to this research, the complete Reddit corpus was not available.[4] In total we crawled 122 thousand posts with 54 million words with an average post length of 71 words for the 1.3 million words sample, and 72 words for the 4.3 million words sample, respectively with standard deviations of 61.7 and 67.8. We show an example Reddit story here:

23

Bad mistake , reasonable doubt , or both ?
So I've been working at this restaurant for the weekend and I was taken aback by the disorganization and bad management .
I don't know my work schedule for the next week ( If I even to continue to work there b/c im on internship in a couple of weeks ) , don't know my exact duties , etc .
The owner and head chef tell me two different things .
I have no idea what's going on .
I made a hasty decision to ask for pay from the owner because of this .
I thought that he could make up my hours or worse not even pay me because there is no system of recording hours or a system of anything really haha .
Once I asked the question and got a sour look I realized I should've asked in a different way ( ex . how does the payment method work ? ) .
Do you guys think it was a mistake , case of reasonable doubt , or both ?
Also , should I continue working there after this situation ?

# 4. Experiments

We applied topic modeling, text classification, and coherence tests to the aforementioned data sets in order to compare them.

24

## 4.1. Text Classification Experiments

As a first analysis of the text collection, we set out to train machine-learning classifiers to distinguish dream reports from personal stories automatically. Both the extent to which the classifiers succeed, and the features they use to make their decision, can lead to insights in the differences between dream reports and other narratives. Terms or groups of terms that are identified by classifiers as strong indications that a text is a dream report or not are apparently typical for their respective class.

25

In text classification, a machine learning classifier is fed with labeled documents from which it learns to model the characteristics of the given labels. Its labeling performance is tested by applying the classifier to a held-out set of documents. For this experiment, we used the sets of 4.3 million words for both the dream data and contrasting data.[5]

26

We tokenized all documents with the Stanford Tokenizer.[6] The word tokens were standardized to lowercase. We extracted word unigrams, bigrams, and trigrams as features. To avoid bias from explicit markers of dream reports, we removed any features that contained one of the following words: dream, dreamer, dreamt, dreamed, dreams, awake, awaken, woke.

27

We compared the performance of three different classification algorithms: Support Vector Machines (SVM), Naive Bayes, and Balanced Winnow. We used the libsvm [Hsu et al. 2003] implementation of SVM, with linear kernel and setting the C parameter to 1.0. We applied Naive Bayes by using the Multinomial Naive Bayes implementation in Scikit-learn [Pedregosa et al. 2011].[7] For Balanced Winnow, we made use of the Linguistic Classification System [Koster et al. 2003]. The α and β parameters were set to 1.05 and 0.95 respectively. The major threshold (Θ+) and the minor threshold (Θ−) were set to 2.5 and 0.5. The number of iterations was bound to one.

28

We evaluated the performance of the three approaches by means of ten-fold cross-validation. To avoid author bias, the reports by the same author were kept together in either the test set or the train set during each fold. During each training

29

phase, the 7,500 most frequent features were selected and presented as binary values.

The classification results, micro-averaged over examples, are given in Table 1. All three approaches yield a precision and recall of 0.97, which indicates that the dream and non-dream reports can easily be distinguished with a small remaining margin of error. Table 1 also displays the exact number of documents that were correctly classified. The Balanced Winnow classifier has a slightly higher number of correct classifications than Naive Bayes and SVM.

| Approach | Prec | Recall | F1 | TPR | FPR | AUC | # correct |
|---|---|---|---|---|---|---|---|
| SVM | 0.97 | 0.97 | 0.97 | 0.97 | 0.03 | 0.97 | 38,225 |
| Balanced Winnow | 0.97 | 0.97 | 0.97 | 0.97 | 0.03 | 0.97 | 38,334 |
| Naive Bayes | 0.97 | 0.97 | 0.97 | 0.97 | 0.03 | 0.97 | 38,229 |

**Table 1.** Micro-averaged performance of three classifiers on distinguishing dream reports from reports of real-world events on the 4.3M words corpus (39,480 documents in total). TPR = True Positive Rate. FPR = False Positive Rate. AUC = Area Under the Curve, # correct = number of correctly labeled documents.

The Balanced Winnow classifier returns an interpretable model of the features that the classifier used internally to make its predictions. Upon analysis of the 30 most indicative features of the dream and non-dream classes, we obtained the following insights about the two types of texts:

- Dream reports are characterized by words that convey uncertainty and retrieval from memory: somebody (rank 3), remember (rank 5), somewhere (rank 12) and recall (rank 17);
- Another category of features that have a high rank in dream reports are references to a space or situation: setting (rank 1), riding (rank 8), building (rank 16), swimming (rank 23), table (rank 25) and room (rank 30);
- In contrast to dream reports, personal stories contain indications of specific points in time: 2014 (rank 4), today (rank 9), tonight (rank 19), yesterday (rank 21), day (rank 23) and months (rank 28);
- Personal stories are also distinguished by conversational utterances, such as ':)' (rank 2), please (rank 17), '?' (rank 20) and thanks (rank 27).

## 4.2. Topic modeling

To discover what type of topics are typical for dream reports, we employed an unsupervised method that is currently popular for discovering latent themes or topics in large document collections. Latent Dirichlet Allocation (LDA) [Blei 2012] is a probabilistic generative algorithm that aims to give a broad overview of the topics that occur in a collection of documents. Topics are defined as a distribution over a fixed vocabulary (in practice, a topic consists of a set of semantically related words). Topic modeling is an unsupervised process solely based on word occurrences in documents. LDA assumes that documents are created based on an underlying topic distribution and each document is generated from a mixture of these topics that each have a different proportion in the document. LDA uses an iterative process to estimate this underlying distribution based on the observed words in the text.

We ran experiments with LDA on the full DreamBank sample of 22,046 dreams. We filtered the dream texts to exclude all function words and punctuation marks and only kept those content words that were automatically part-of-speech tagged by the Stanford parser as nouns, verbs and adjectives. All words have been converted to lower case. Such explicit filtering step ensures that the generated LDA topics contain only content words.

For these experiments we use the LDA implementation provided in the Mallet toolkit [McCallum 2002]. We ran LDA with 2,000 iterations with Gibbs sampling and 50 topics. The produced LDA model was used to annotate each document with its most relevant topics; namely those topics that cover at least 10 percent of the document. Documents have three such topics on average.

Setting the number of topics parameter is a rather arbitrary choice. We ran experiments with 100, 200 and 400 topics as well and studied the output. When raising the value of this parameter, more fine-grained topic descriptions are produced. These detailed topics are still understandable and coherent topics, but, as can be expected, they tend to

have a lower coverage in the document. As we aim to look at significant differences between topic distributions in different sample sets and to compute g-tests (log-likelihood tests) [Rayson and Garside 2000], we choose to keep the number of topics fixed at 50.

### 4.2.1. LDA on the DreamBank

LDA can give surprising insights in the data. We applied LDA to the full DreamBank set of dreams and we present a random sample of these topics in Table 2. The number in each row denotes the topic number and does not express a ranking or weight. Certain topics express a specific script or frame; in the first three topics in Table 2 we see "purchasing", "using the bathroom" and "school life". Other topics express a setting such as "inside the house" in topic 42, and the "outdoor" setting in topic 5. It is also remarkable to see how narrative verbs are clustered together in present tense in topic 35 and in past tense in topic 48. These verbs are commonly used in action and event descriptions (do, say, see). These automatically generated topics are a clear support for the *continuity hypothesis* [Domhoff and Hall 1996] as they reflect daily life events, characters and settings.

36

| 44 | money pay get give buy bank bill machine change |
| 37 | bathroom water toilet shower use clean bath floor sink |
| 25 | class school teacher students high test room classroom college |
| 42 | room door house see window open apartment go living |
| 5 | road hill tree see walking snow mountain trees people |
| 28 | love feel kiss make happy want man other hug |
| 35 | say says do see go man woman comes get |
| 48 | said did went came got told started saw looked asked |

**Table 2.** Examples from the topics generated on the DreamBank sample.

In a next step we zoom in on two comparable dream sets of men and women to study the differences in topics between these groups. We use the normative male and female dream sample present in the DreamBank (abbreviated to *Hall/VdC Norms*) based on the older work of Hall and Van de Castle [Hall and Van de Castle 1966].

37

Topics were generated based on the full sample. For each topic we compute whether the topic occurs significantly more or less in *Hall/VdC Norms* male dreams than female dreams. We computed a g-test[8] on the frequency topic counts with $p<0.05$. In Table 3 we show the topics that occurred significantly more in either the male or the female dream sub-sample. Remarkable stereotypical differences can be found in the topics; men tend to dream more about shooting, driving, sex, and games, while women dream more about weddings, fashion, and family.

38

LDA topics have been shown to express semantic coherence. Although there is currently no metric available that could be used to optimize the LDA settings to tune it explicitly towards human judgments, it has been shown that the automatic topic assignment to documents matches human preferences [Chang et al. 2009]. Nonetheless, the human interpretation of automatically generated topics might not always be straightforward such as topics 11 or 39, which both seem to grasp a less clear topic related to dreaming in general.

39

These preliminary results are in line with recent research on differences between male and female in [Domhoff and Schneider 2015], and other work such as [Schredl and Piel 2005] and [Domhoff and Schneider 2008], which states "[that] there are more appearances of tools and cars in men's dreams, more appearances of clothing and household items in women's dreams". The main difference is that the results presented here were obtained unsupervised, and support the current manually found results reported in other papers.

40

| Male | |
|---|---|
| 0 | gun fire men shot shoot man police shooting war deer |
| 5 | road hill tree see walking snow mountain trees people side |
| 11 | dream remember seemed do time being other same feeling something |
| 15 | car driving drive road street get truck going front side |
| 17 | bed room sleep sex sleeping lying bedroom floor naked lay |
| 29 | game playing ball play team basketball football field baseball cards |
| 33 | plane fly flying sky air see airplane land people ground |
| 34 | building floor stairs get go elevator people steps see walking |
| Female | |
| 12 | wedding married john wife getting ring husband george bonita ceremony |
| 14 | things room put stuff box small old boxes take find |
| 21 | wearing white dress black blue dressed clothes red shirt shoes |
| 23 | house mother father home brother family old sister parents children |
| 26 | get go do going trying take want find time know |
| 39 | girl friend dream girls friends remember went dreamed did school |
| 40 | man woman men young other women boy old older small |

**Table 3.** Male (top) and female (bottom) topics in the DreamBank.

### 4.2.2. LDA on dreams versus personal stories

In the next step we combined the dream sample with the Reddit and Prosebox samples into one large collection on which we ran the LDA topic modeling using the same setting of 50 topics. [41]

To investigate what topics occur significantly more with dreams than with personal stories, we took a random sample of 2,000 dreams and 2,000 stories and computed a g-test to check whether there were significant differences in the topic distributions. In this sample of 4,000 documents we found that 42 of the 50 topics occur significantly more or significantly less with either dreams or personal stories. This shows that there are clear differences between the two sets; more so than between the male and female dreams. In Table 4 we show the top five most significantly different topics for dreams and stories. [42]

Topic 28 is typical for what we expect to be a dream description, mentioning words such as "dream", "remember" and "seemed". We observe an "inside the house" setting description (topic 23) and an aquatic setting description (topic 1). The other two topics express related narrative description verbs in present and paste tense. [43]

For the personal stories we observe two topics that are directly linked to the Reddit categories that were included in the sample, namely "anxiety" and "relationships". Topic 45 expresses conversational internet language including profanities and abbreviated verb forms. Topic 24 consists mostly of time expressions. [44]

There is some overlap in the most important words in the topic word lists. The term "get" occurs in the top words lists of four of the five most significant personal stories topics. The terms "see" and "saw" occur in respectively three and two of the significant dream topics. [45]

| Dreams | |
|---|---|
| 23 | room house door see go floor open stairs window apartment |
| 40 | saw came said went looked got ran walked horse did |
| 1 | water see boat pool river swimming lake beach go people |
| 28 | dream remember seemed girl man boy came saw dreamed being |
| 13 | see go says say man woman get look comes walk |
| **Non-Dreams** | |
| 26 | do time get things think going know something make other |
| 45 | fucking shit fuck do get im know day got dont |
| 30 | feel do life help depression anxiety get know feeling want |
| 22 | relationship do want feel other months boyfriend tl girlfriend friends |
| 24 | day last today work get going week night got time |

**Table 4.** Five most significant topics for dreams (top) and non-dreams (bottom).

## 4.3. Bizarreness as dream characteristic

When people are asked what is typical about dreams, they will often mention weirdness as a typical property of dreams. This might be due to the fact that many dreams are forgotten the next morning while weird or impressive dreams tend to stick in people's memory [Bulkeley and Hartmann 2011]. This recollection could be attributed to the bizarreness effect, the inclination to remember bizarre things better than ordinary things [McDaniel et al. 1995]. Domhoff shows in [Domhoff 2007] that bizarreness does occur in dreams, but it is not as manifest as people tend to believe.

Bizarreness can emerge in different forms in dreams. The most prominent type of oddity seems to be caused by the discontinuity of events and sudden switches of scenes. In the study of Reinsel et al. [Reinsel et al. 1992] bizarreness was measured by counting three different type of occurrences: discontinuous events, improbable combinations, and improbable entities. Discontinuous events were found to be the most contributing factor in bizarreness (around 60% of the counts) while improbable entities were much less present (only around 8%). This is in line with previous studies on large volumes of dream reports that show that the amount of bizarreness attributed to strange entities is relatively low; most characters in dreams are known persons [Domhoff and Schneider 2008]. Domhoff [Domhoff 2003, 132] investigated metamorphosis, a specific type of discontinuous events, via a search in DreamBank.net and found only 50 mentions of metamorphosis in the whole DreamBank. Transformations turn out to be highly infrequent in dreams. [Schweickert and Xi 2010] study metamorphosis as a typical dream phenomenon and focus on the relation between change in form to change in inner states. No evidence was found that form change is connected to a change in mental state. This was a small-scale study on a set of 65 dreams from 21 persons.

Interestingly, bizarre thoughts do not only occur in our sleep but can also occur when awake. It has been shown that people in a relaxed undisturbed awake condition produce dream-like reports when asked to recall that most recent thoughts in the same way subjects are asked after being awakened [Reinsel et al. 1992]. These awake fantasies are very similar to dream reports, including bizarreness.

In our study we aimed at using a quantitative approximation of bizarreness and applying this metric to the DreamBank texts. We focus on the discontinuity of events in dreams and try to quantify this by looking at textual coherence in the dream reports. We hypothesize that dreams are less coherent in their discourse structure than personal stories. We measure two different aspects of discourse structure, namely discourse marker frequency and entity-based text coherence, using the smaller balanced sample sets of 1.3 million words.

Discourse analysis is a broad and multi-disciplinary field that studies language in use beyond the sentence level [Trappes-Lomax 2004]. Automatic discourse parsing is still in its early development phase, as was illustrated by 2015's CoNLL shared task [Xue et al. 2015] on shallow discourse parsing where the best system achieved a overall F-score of

24%.

In this initial experiment we only focus on discourse marker occurrences and measure whether there is a difference between discourse marker frequency in the dream data and the personal stories. Discourse markers are words or phrases that explicitly signal discourse structure and describe how two sentences or phrases are related to each other. For example, *for example* indicates that the current sentence exemplifies something that was mentioned in the previous sentence. Typical discourse markers are *but*, *since*, *while*, *even though*, and *because*.

We used a list of 60 markers[9] based on annotations from the Penn Discourse Treebank [Prasad et al. 2008] that was used in the CoNLL shared task. In total, about 40 thousand occurrences of these discourse connectives were counted in the DreamBank data, and about 50 thousand in both Reddit and Prosebox, which means that there are about 20% fewer discourse markers counted in the DreamBank data than in the contrasting data. In Figure 1 we show the distribution of 28 of the 60 markers in the balanced DreamBank, Reddit and Prosebox data sets, having a frequency of occurrence over 250. For most markers we can observe similar distributions, or a slightly lower count for the dreams. One marker however occurs substantially more often with the dreams than with Prosebox or Reddit: "then". This is a typical discourse connective that is used in sequential narration. Conditional discourse markers like "if" and "if then" and causal markers like "because", "so" and "since" occur less frequent in the dream sample.
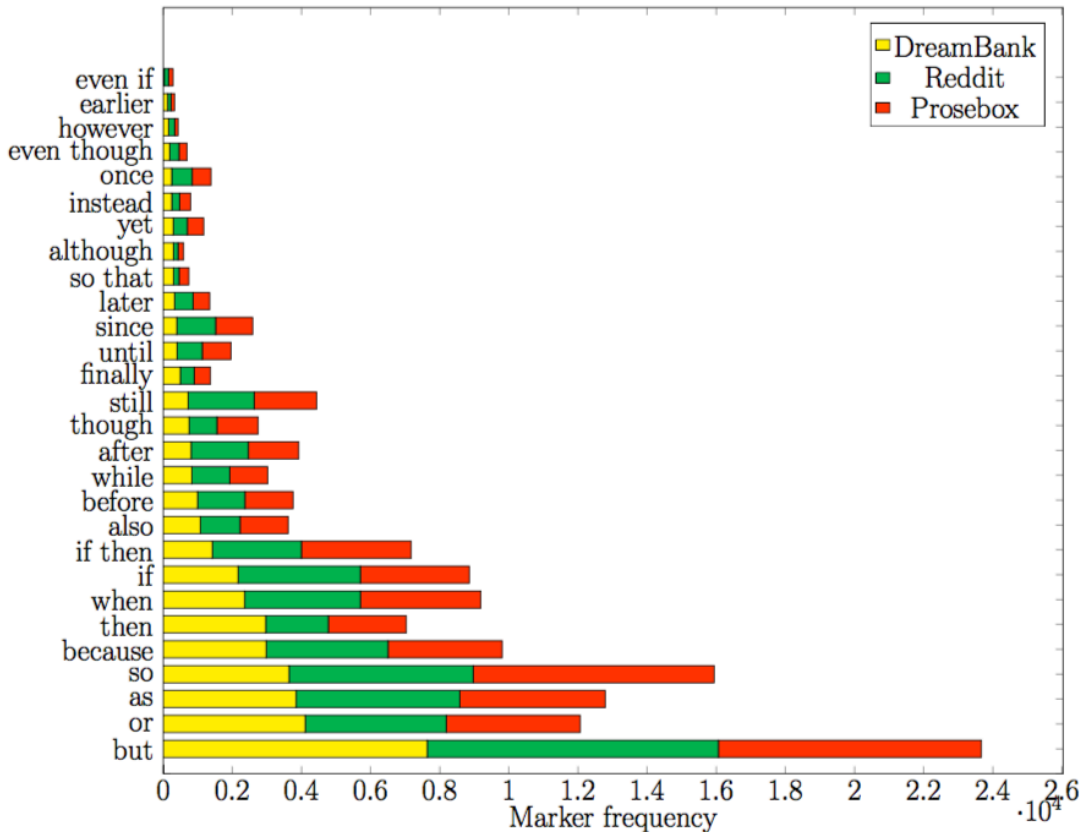
**Figure 1.** Frequency of discourse markers per dataset and their total number. Discourse markers used have a frequency of more than 250 in the Penn Discourse Treebank.

In a second experiment we study entity-based coherence. Mentioned entities and chains of referring expressions in a text are core indicators of text coherence. We assume that discontinuity in dreams is expressed in sudden shifts in scenes and events, and we expect that these are linked to shifts in mentioned entities. On the basis of this assumption, we tried to measure discourse coherence by applying an existing automatic model to detect entity-based coherence.

We used the Brown Coherence Toolkit v1.0 [Elsner and Charniak 2008]. The authors of this toolkit present an extension of the entity-grid coherence model proposed by Barzilay and Lapata [Barzilay and Lapata 2008]. An entity grid

represents the entity mentions in a document in a textual matrix where each row represents an entity and the column represent the syntactic roles of the entities (subject, object, other). This matrix is used to predict which role each entity will have in the next sentence.

To detect the entities in the text we used the extended entity grid based on the Wall Street Journal corpus that was automatically pre-processed with OpenNLP software, available in the Brown Coherence Toolkit.

<div style="text-align: right">55</div>

We applied the model to each of the balanced dream and personal stories data sets and measure its performance with a binary discrimination test as was previously done in the work of Elsner and Charniak [Elsner and Charniak 2008]. The binary discrimination test tests the model's ability to distinguish between a human-authored document in its original order, and a random permutation of that document. The test reads any number of documents and performs the test on each one, using 20 random permutations.

<div style="text-align: right">56</div>

The results of this test are shown in Table 5. All achieved results are substantially lower than the scores reported by Elsner and Charniak, who report an F-score of 86%, when training on *Wall Street Journal* (WSJ) newspaper text and testing on a held-out set from the same corpus. As WSJ consist of financial news paper articles, we can expect a drop in performance when switching to a completely different textual genre of dreams and personal stories. Nevertheless, the scores in Table 5 indeed suggest that the dream text is less coherent than more formal edited text in terms of coherence relations, but also as compared to Prosebox and Reddit, which are remarkably similar.

<div style="text-align: right">57</div>

| Dataset | Accuracy | F-score |
|---|---|---|
| DreamBank | 0.23 | 0.32 |
| Prosebox | 0.37 | 0.42 |
| Reddit | 0.37 | 0.43 |

**Table 5.** Results from the entity-based coherence model as evaluated by a binary discrimination test.

# 5. Discussion

We presented three automatic text analysis studies of dream reports. First, we performed a supervised text classification experiment to see how easy or hard it is to distinguish dream reports from texts that are closely related in both content and structure, namely true personal stories. We applied three different text classification algorithms to the same task; they all succeeded in labeling the documents with a near-perfect precision. Differentiating between dreams and personal stories turned out to be an easy task. The analysis of the features used by the Balanced Winnow classifier shows that expressions of uncertainty, setting descriptions and narrative verbs are typical for dreams, while time expressions and conversational expressions are typical for the personal stories.

<div style="text-align: right">58</div>

In the second study we aimed to explore the general topics that are present in the full DreamBank. We applied LDA topic modeling to the DreamBank to study the main themes in dreams. The results mostly showed topics describing everyday activities, settings, and characters. This unsupervised method signaled the same differences as the text classification experiments between dreams and stories: setting descriptions and uncertainty expressions are typical for dreams while time expressions and conversational expressions occur more often in stories. In our exploratory study on discontinuity in dreams, we observed that dream reports indeed use less discourse markers and have a lower entity-based textual coherence. With these experiments we are only just scratching the surface of doing automatic discourse analysis but we feel that these preliminary experiments are a starting point for further analyses in this direction.

<div style="text-align: right">59</div>

Even though our experiments have shown some interesting and consistent findings about the typical differences between dreams and stories, we need to be careful with our conclusions. The fact that the text classifiers obtained such high scores and the topics were significantly differently distributed over the samples, could indicate that the contrasting data sample was not as representative as we had hoped for. The emerging topic about "anxiety" for example shows that this subreddit had a substantial influence on the results. The importance of conversational expressions to distinguish personal stories in the classification experiment also points to a bias of the source platforms. We suspect that a more

<div style="text-align: right">60</div>

careful selection over a much larger set of personal stories, and perhaps an additional check to filter out characteristic internet language is needed to create automatic models that focus on the more subtle differences between reported dreams and personal experiences from real life.

As a next step, we plan experiments on another sample of personal stories and dreams to investigate the effect of the sample representativeness. We also aim to collaborate with dream analysis experts to work towards better interpretations of the results that we found and to explore further research questions in the area of dream analysis. [61]

We also believe that an in-depth study of the narrative mechanisms in dream descriptions could be a fruitful path for future research. The overview of the distribution of discourse markers presented in Figure 1 begs for further analysis. Furthermore, we would like to investigate the applicability of the Labovian Model of narrative [Labov and Waletzky 1967] to dream descriptions. We would expect that dreams in general contain the narrative units *orientation* and *complicating action* but that the *coda* and *resolution* are often absent. [62]

Furthermore, we are interested in the question whether humans can just as easily distinguish between dream descriptions and true stories as the text classifier could. We are currently working on building an online human judgment task to investigate this question. [63]

# Acknowledgments

## Notes

[1] We used langid.py version 1.1.5 (github hash: e801bf8, accessible at http://git.io/vcc2Z).

[2] We used twokenize.py, which is part of twitter_nlp (github hash 27c8190, accessible at http://git.io/vccyu).

[3] https://www.prosebox.net/

[4] See https://redd.it/3bxlg7 for a dataset with all 1.7 billion publically available Reddit posts.

[5] We ran the same experiments on the small sets too and found virtually the same results.

[6] http://nlp.stanford.edu/software/tokenizer.shtml

[7] http://scikit-learn.org/stable/modules/naive_bayes.html

[8] We used the g-test implementation written by Pete Hurd, available at http://www.psych.ualberta.ca/~phurd/cruft/g.test.r

[9] We excluded "and" as discourse marker in this experiment due to its ambiguity as conjunction marker and high frequency.

## Works Cited

**Barzilay and Lapata 2008** Barzilay, R. and Lapata, M. (2008). "Modeling local coherence: An entity-based approach." *Computational Linguistics*, 34(1), pp. 1–34.

**Black and Green 1992** Black, J. and Green, A. (1992). *Gods, Demons and Symbols*. University of Texas press.

**Blei 2012** Blei, D. (2012). "Probabilistic topic models." *Communications of the ACM*, 55(4), pp. 77–84.

**Bulkeley 2009** Bulkeley, K. (2009). "Seeking patterns in dream content: A systematic approach to word searches." *Consciousness and cognition*, 18(4), pp. 905–916.

**Bulkeley 2014** Bulkeley, K. (2014). "Digital dream analysis: A revised method." *Consciousness and Cognition*, 29, pp. 159 – 170.

**Bulkeley and Hartmann 2011** Bulkeley, K. and Hartmann, E. (2011). "Big dreams: An analysis using central image

intensity, content analysis, and word searches." *Dreaming*, 21(3), pp. 157.

**Chang et al. 2009** Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., and Blei, D. (2009). "Reading tea leaves: How humans interpret topic models." In *Advances in neural information processing systems*, pp. 288–296.

**Domhoff 2003** Domhoff, G. (2003). *The scientific study of dreams: Neural networks, cognitive development, and content analysis.* American Psychological Association.

**Domhoff 2006** Domhoff, G. (2006). "Barb Sanders: our best case study to date, and one that can be built upon." http://www2.ucsc.edu/dreams/Findings/barb_sanders.html

**Domhoff 2007** Domhoff, G. (2007). "Realistic simulation and bizarreness in dream content: Past findings and suggestions for future research." *The New Science of Dreaming: Content, Recall, and Personality Characteristics*, 2, pp. 1–27.

**Domhoff and Hall 1996** Domhoff, G. and Hall, C. (1996). *Finding meaning in dreams: A quantitative approach*. Springer.

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

**Domhoff and Schneider 2015** Domhoff, G. and Schneider, A. (2015). "Correcting for multiple comparisons in studies of dream content: A statistical addition to the Hall/Van de Castle coding system." *Dreaming*, 25(1), pp. 59.

**Elsner and Charniak 2008** Elsner, M. and Charniak, E. (2008). "Coreference-inspired coherence modeling." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pp. 41–44. Association for Computational Linguistics.

**Frantova and Bergler 2009** Frantova, E. and Bergler, S. (2009). "Automatic emotion annotation of dream diaries." In *Proceedings of the Analyzing Social Media to Represent Collective Knowledge Workshop at K-CAP 2009, The Fifth International Conference on Knowledge Capture*.

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

**Hsu et al. 2003** Hsu, C., Chang, C., Lin, C., et al. (2003). *A Practical Guide to Support Vector Classification.* Department of Computer Science, National Taiwan University

**Hurovitz et al. 1999** Hurovitz, C., Dunn, S., Domhoff, G., and Fiss, H. (1999). "The dreams of blind men and women: A replication and extension of previous findings." *Dreaming*, 9(2-3), pp. 183–193.

**Jung and Shamdasani 2010** Jung, C. G. and Shamdasani, S. (2010). *Dreams:(From Volumes 4, 8, 12, and 16 of the Collected Works of CG Jung)(New in Paper)*, volume 20. Princeton University Press.

**Kahan and LaBerge 2011** Kahan, T. and LaBerge, S. (2011). "Dreaming and waking: Similarities and differences revisited." *Consciousness and Cognition*, 20(3), pp. 494 – 514.

**Kilroe 2000** Kilroe, P. (2000). "The dream as text, the dream as narrative." *Dreaming*, 10(3), pp. 125–137.

**Koster et al. 2003** Koster, C., Seutter, M., and Beney, J. (2003). "Multi-classification of patent applications with Winnow." In *Perspectives of System Informatics*, pp. 546–555. Springer.

**Labov 2006** Labov, W. (2006). "Narrative pre-construction." *Narrative inquiry*, 16(1) , pp. 37–45.

**Labov 2008** Labov, W. (2008). "Oral narratives of personal experience." *Cambridge encyclopedia of the language sciences*.

**Labov and Waletzky 1967** Labov, W. and Waletzky, J. (1967). "Narrative analysis: Oral versions of personal experience." *Essays on the Verbal and Visual Arts*, pp. 12–44.

**Lui and Baldwin 2012** Lui, M. and Baldwin, T. (2012). "langid. py: An off-the-shelf language identification tool." In *Proceedings of the ACL 2012 system demonstrations*, pp. 25–30. Association for Computational Linguistics.

**Matwin et al. 2010** Matwin, S., De Koninck, J., Razavi, A., and Amini, R. (2010). "Classification of Dreams using Machine Learning." In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI)*, volume 215, pp. 169–174, Lisbon, Portugal. IOS Press.

**McCallum 2002** McCallum, A. (2002). MALLET: A machine learning for language toolkit. http://mallet.cs.umass.edu.

**McDaniel et al. 1995** McDaniel, M., Einstein, G., DeLosh, E., May, C., and Brady, P. (1995). "The bizarreness effect: it's not

surprising, it's complex." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), p. 422.

**Montangero 2012** Montangero, J. (2012). "Dreams are narrative simulations of autobiographical episodes, not stories or scripts: A review." *Dreaming*, 22(3), p. 157.

**Pedregosa et al. 2011** Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12, pp. 2825–2830.

**Prasad et al. 2008** Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., and Webber, B. (2008). "The Penn discourse treebank 2.0." In *Proceedings of the sixth International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA).

**Rayson and Garside 2000** Rayson, P. and Garside, R. (2000). "Comparing corpora using frequency profiling." In *Proceedings of the workshop on Comparing Corpora*, pp. 1–6. Association for Computational Linguistics.

**Razavi et al. 2014** Razavi, A., Matwin, S., De Koninck, J., and Amini, R. (2014). "Dream sentiment analysis using second order soft co-occurrences (SOSCO) and time course representations." *Journal of Intelligent Information Systems*, 42(3), pp. 393–413.

**Reinsel et al. 1992** Reinsel, R., Antrobus, J., and Wollman, M. (1992). "Bizarreness in Dreams and Waking Fantasy." *The neuropsychology of sleep and dreaming*, pp. 157–181.

**Schredl and Piel 2005** Schredl, M. and Piel, E. (2005). "Gender differences in dreaming: Are they stable over time?" *Personality and Individual Differences*, 39(2) , pp. 309–316.

**Schweickert and Xi 2010** Schweickert, R. and Xi, Z. (2010). "Metamorphosed characters in dreams: Constraints of conceptual structure and amount of theory of mind." *Cognitive Science*, 34(4) , pp. 665–684.

**Trappes-Lomax 2004** Trappes-Lomax, H. (2004). "Discourse analysis." *The handbook of applied linguistics*, pp. 133–164.

**Xue et al. 2015** Xue, N., Ng, H., Pradhan, S., Prasad, R., Bryant, C., and Rutherford, A. (2015). "The CoNLL-2015 shared task on shallow discourse parsing." In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning - Shared Task*, pp. 1–16, Beijing, China. Association for Computational Linguistics.