# You Are Not Feeling Sleepy: A Computational Analysis of Hypnotic Linguistics

Benjamin Shapiro (benj@berkeley.edu)
Department of Cognitive Science, UC Berkeley

**Abstract**

This computational study attempts to quantify what it means for language to be 'hypnotic' by training a naive Bayes classifier to distinguish hypnotic scripts (documents used for linguistic reference in conducting hypnosis sessions) from "normal" English speech or writing with no hypnotic content. The results show that "hypnotic" speech contain certain distinct key linguistic elements that set it far apart from standard english use, and that computational analysis of hypnotic scripts can be useful in understanding and formalizing the intricacies of hypnosis.

**Keywords:** natural language processing; hypnosis; linguistics; Bayes

## Introduction

Hypnosis is a scientifically ill-defined process. The processes by and through which it occurs have been minimally researched (due to the largely subjective and vague nature of hypnotic experience), with some success; this research has focused mainly on observation of hypnotized subjects, sometimes neurologically (Kihlstrom, 2007). However, the truth behind the neural and cognitive mechanisms of hypnosis is still largely up for debate (Kihlstrom, 2008; Kihlstrom, 2007), and no computational study has been done on hypnotic phenomena thus far.

For the purposes of this paper, hypnosis or hypnotic process shall be defined as an instance in which a subject is, of their own volition, in trance. Trance is usually loosely classified as an "altered state of consciousness" (Kihlstrom, 2007) in which the subject is highly susceptible to suggestion and direction. The process by which a subject is introduced into trance is called the *hypnotic induction* (Gafner and Benson, 2000). Although hypnotic induction does not always require speech, for the purposes of this study, we will focus on examples of vocal hypnotic induction in an effort to isolate linguistic content from any other variables which might be contributing to the subject's journey into trance. Specifically, this study will be focusing on analysis of hypnotic scripts: documents that hypnotists sometimes read from as a way to conduct their hypnotic induction.

In an effort to answer the obvious question, "What is hypnosis, really?", a naive Bayes model (Jurafsky, Martin, 2009) was used to analyze the linguistic content of hypnotic scripts, with the hope of answering a simpler, but useful question: "What causes hypnosis?" The goal of this analysis is to begin building a formal, computational and descriptive model of which linguistic variables play a part in inducing a state of trance. Such a model would be useful in understanding how hypnosis works, and ultimately how the mind responds to elements of hypnotic language. Results showed that many heuristics, or "rules of thumb" for hypnotic inductions have solid computational foundations in hypnotic scripts, but there are also some unexpected intricacies which show that further study of hypnosis from a computational perspective could afford useful insight into hypnotic phenomena.

## Background

In the following analysis, we explore whether the linguistic trends of hypnotic scripts are consistent with theoretical claims of what language should be used in the process of hypnotic induction. To frame this analysis, we first observe several key theoretical claims in the available literature on hypnosis and hypnotic language. Hypnosis, whether performed on stage or in a research context, appears to be an extremely complex process, and trance can only be induced successfully with the help of certain guidelines with regards to what to say during a hypnotic induction, and how to say it. Although the content of a hypnotic script will vary depending on its source, hypnotists will tend to consistently follow certain guidelines with regard to language choice. It will be useful for the reader to keep in mind that many of these rules of thumb will vary in their usage over different methods, hypnotists and contexts, and are also only a few of many intricacies in correctly performing hypnotic induction (Gafner and Benson, 2000).

Firstly, hypnosis is often conducted with the use of certain language that promotes relaxation (Gafner and Benson, 2000; Hammond, 1990). For many readers, the classic movie phrase "You are feeling veeeerryy sleeeeeepy" will come to mind. According to much hypnosis theory developed by stage hypnotists and hypnotherapists alike, this may not be so far off (save for the exaggerated vocal tonalities), but should be accompanied with other key words such as "sleep", "relax," "calm," "peaceful," "tired," and so forth. "Sleep" has been theorized to be an especially powerful key word for hypnotic inductions; in the case of rapid inductions (or "instant" inductions, wherein a hypnotist may induce a state of trance in as little time as ten seconds), "sleep" is often the only key word uttered after a short, standard pre-trance talk detailing the effects and procedure of hypnosis. In fact, the famous "8 Word" or "Hand Drop" induction is especially popular among novices and experts alike when inducing rapid trance; the only words uttered during the induction process are "Press on my hand. Close your eyes. SLEEP!" Interestingly enough, our analysis in the upcoming sections will show that "sleep" as a keyword is not actually quite as important to hypnotic language as we might expect, given this background.

Further, negation and negative suggestions are discouraged (Gafner and Benson, 2000; Hammond, 1990). There is a psychologically simple reason for this. When asked to "not imagine a large balloon," subjects will first imagine a balloon, then attempt to "cross out" the image from their minds. Therefore, when a hypnotist wishes to suggest to a subject that they are unable to move their arm, they are more likely to say “your arm is stiff” than they are to say “your arm does not bend.” This makes the suggestion more powerful to the subject than if the hypnotist had first evoked a contradictory concept within the subject’s mind.

Lastly (for the purposes of this analysis), hypnotic inductions often evoke imaginative images (Kihlstrom, 2007; Gafner and Benson, 2000; Hammond, 1990; Hewitt, 2003). The goal of hypnotic language is to focus the subject’s attention away from their present context, mentally or physically, in other words, to create an “out of body” experience. For this reason, hypnotists tend to ask subjects to direct their attention to intangible emotions, feelings or perceptions, in an attempt to relax them and distract them from awareness of reality. This will often involve inductions with the use of the subject’s hands, and the consistent use of dissociation terms like “those,” “that,” or “there,” rather than “these,” “this,” or “here,” but more importantly, hypnotic language must be descriptive and evocative.

With these and many other concerns in mind, hypnotists write hypnotic scripts to assist them in conducting their hypnosis session. While the variety of scripts available in hypnosis literature is wide, this study will mostly be an analysis of general hypnotic language, without regard for topical hypnosis scripts such as those whose purpose is to remove a smoking addiction or assist with weight loss. In other words, the text we are concerned with is in those hypnotic scripts whose main purpose is simply to achieve a state of trance. As such, the majority of the language should be governed almost entirely by the above rules.

## Model

A naive Bayes model was employed for this analysis. Documents were divided into two categories, “hypnotic” and “not hypnotic.” A naive Bayes classifier with words as features was trained with these categories in an effort to distinguish hypnotic from non-hypnotic lexical variables; this model does not provide us with any information about sentence structure, but it provides us with statistics pertaining to the lexical content of the documents, which can give some hints about the underlying structure of documents in each category (Jurafsky, Martin, 2009). The naive Bayes document classifier calculates the probability of a document by assuming conditional independence of every word given a document category. While not necessarily an entirely realistic assumption, lack of conditional independence would make our problem intractable. Fortunately, this assumption still leaves us with enough information to infer general document content and structure.

Given our model, our variables are as follows. $P(D|C)$ is the probability of a document given a document category. Given this likelihood and a prior $P(C)$, we are able to calculate the posterior probability that a document D is in category C ($P(D|C)$) with the simple application of Bayes’ rule:

$$P(C|D) \propto P(D|C) * P(C) \tag{1}$$

According to the naive Bayes model, we are able to assume that words are conditionally independent given a category, and can therefore generalize $P(D|C)$ to the following computation:

$$P(D|C) = P(w_1|C) * P(w_2|C) * ... * P(w_n|C) \tag{2}$$

where $w_i$ is the $i^{th}$ word in the document. Due to the inevitably large number of words present in each document and (as an end result) our entire corpus, log likelihoods were utilized in order to keep the statistics manageable and comparable, such that:

$$P(w|C) = \log(nWordInClass) - \log(nWordsInCategory) \tag{3}$$

where *nWordInClass* is the total number of appearances of word $w$ in category $C$, and *nWordsInCategory* is the total number of appearances of all words in category $C$.

Additionally, the model employed the concept of “stop words” in order to improve the quality of the statistical results. The list of stop words, which is publicly available (Brahaj, 2009) and was modified slightly to account for words we thought may be statistically significant, includes words such as “the,” “a,” “and,” and “or,” which were excluded from each document model. Simple Laplace smoothing was applied to word counts in order to ensure that words would never be assigned a probability of 0; all words not found in the corpus were counted as having been found a single time. This kind of smoothing is extremely simplistic in the sense that it is unrealistic to assume that because it is not in our (somewhat limited) corpus, it would only appear once in an accurate model of the human language. However, for the purposes of building document category models, this is good enough (Jurafsky, Martin, 2009).

In choosing content with which to train and test the model, the corpus for each category was built as follows. The hypnotic category contained scripts pulled from websites on hypnosis, scripts found in books written by stage hypnotists and scripts used in tests for hypnosis susceptibility by institutes such as Stanford University and Arizona University (Weitzenoffer, 1962; Kihlstrom, 1994). The ”not hypnotic” category contained random short stories pulled from various short story databases on the internet; these were chosen such that their lengths were approximately consistent with documents in the hypnotic scripts category. Short stories were specifically chosen for this category because they constitute a large of range of topics and styles of English. For the purposes of generating a model of an“ideal” hypnotic induction, it may have been more effective to limit ourselves to professional and research quality hypnotic scripts, but such documents are few and far between.

## Results

Firstly, the classifier itself distinguished between documents from the hypnotic category and the short story category very accurately. Even when trained with a single document in each

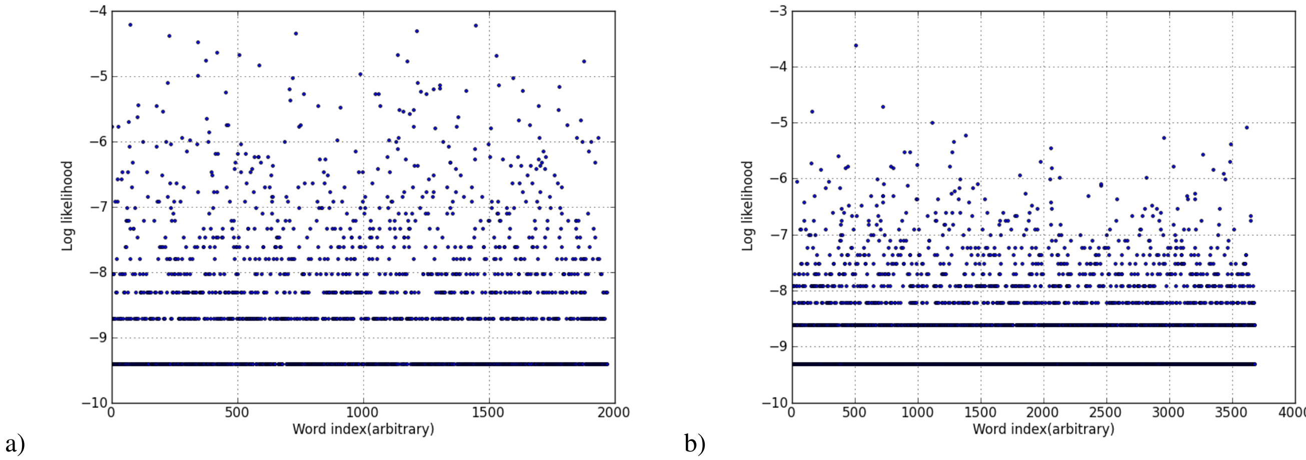


**Figure 1:** These graphs display the log likelihoods returned by the model. These are the probabilities of each word given a category (namely, $\log(P(w_i|C))$), for the hypnotic category on the left (1a) and short story on the right (1b). 1a) exhibits a large distribution of words which have high likelihoods, indicating that the documents in this class are probably more structured and tend to centre around specific key terms, whereas 1b exhibits a much larger concentration of words which have very small likelihoods, and does not exhibit nearly as much tendency toward key terms. Additionally, the short story distribution exhibits a much larger range of unique terms (as is apparent from the proportion of the x axis), which is to be expected from a random sample of unstructured literature.

category, the results of classification with a document of either type were always accurate. In other words, the distinctions between documents of each category are so extreme that even a single sample of each type of document is enough to make accurate future predictions of document category. It is therefore reasonable to examine the data that the model has learned (especially with a larger set of training documents) and use them to make judgements about rules concerning hypnotic language.

Figure 1 shows the statistical results of the model when run with a moderate corpus of documents for each category. The x-axis of each graph is simply the index of any given word (these indexes are assigned arbitrarily; they exist only to keep track of each individual word), while the y-axis is the respective log likelihood of the word: that is, the natural log of the probability of that word given the category that the graph represents. Care was taken to ensure that each corpus had approximately the same amount of data, so the total number of words for each category can be assumed to be constant.

Notably, there are several qualitative differences to the categories of hypnotic and non-hypnotic documents respectively, as is shown by the graphs. Firstly, the non-hypnotic category (Figure 1b) contains nearly twice as many unique words as did the hypnotic category (Figure 1a). Naturally, this is to be expected; short stories will tend to have quite a few more distinct words, used without repetition, particularly since these documents do not need to be relevant to each other in any way. However, this also shows that hypnotic documents tend to place much more weight on using the same linguistic variables over and over again, which tells us that a) repetition is an important aspect of hypnotic linguistics and b) there is a set of key terms which get repeated much more often than others in hypnotic scripts, whereas words in short stories tend to be more equally likely. As a result, it can clearly be seen that most of the log likelihoods in Figure 1b cluster just below -9.0; in other words, all points in this area have a probability of near-zero (we will call this area the baseline). The probabilities in this figure are so concentrated around this baseline that there is essentially a solid line there; while one would expect that there would also be a concentration of probability around this baseline in Figure 1a, there is slightly less of such a weight on near-zero probabilities due to the greater amount of semantic consistency throughout the hypnotic corpus. Additionally, we can see that there is a fair amount of weight in hypnotic documents that shifts to the area of probability above the -5.0 mark; this probability concentration does not nearly exist to the same extent among short stories.

A quick check of some of the words that occur with the highest probability in each class yields the results shown in Figure 2, starting with the most probable word at the top, and continuing down in order of decreasing probability.

## Discussion

In light of our statistical results, let us now turn to an analysis of the previously explored theoretical claims about hypnotic language. Firstly, we made note of a tendency towards certain key words that promote relaxation, and the overwhelming power of the term "sleep" in linguistic contexts as well as its marked usage in popular culture depictions of trance. The data in the table above certainly supports the heuristic of relaxing terms; "relax," "relaxed," and "deeper" all fit this profile very well. Additionally, "heavy" is another relaxing term often used by hypnotists with relation to parts of the body, especially the hands, arms, legs or eyelids.(Hammond,

| hypnotic | log likelihood | not hypnotic | log likelihood |
|---|---|---|---|
| relax | -4.21 | said | -3.62 |
| relaxed | -4.23 | old | -4.71 |
| feel | -4.31 | mr | -4.80 |
| eyes | -4.35 | man | -5.00 |
| mind | -4.39 | hand | -5.09 |
| deeper | -4.48 | mrs | -5.23 |
| right | -4.64 | summers | -5.26 |
| that's | -4.68 | white | -5.34 |
| time | -4.68 | like | -5.38 |
| body | -4.68 | did | -5.52 |
| heavy | -4.78 | little | -5.52 |
| hand | -4.78 | wish | -5.57 |
| like | -4.78 | took | -5.59 |

**Figure 2:** The most common words in each category, sorted from most to least common.

1990) Note, however, that the term "sleep" is not even present in these top hypnotic concepts. In fact, it takes a fair amount of digging towards the baseline (near-zero probability words) before "sleep" is discovered to be a higher probability term than most. In fact, while "sleep" does not have a near-zero probability, the data above shows that there are many other terms which are much more prevalent. We can only guess as to why this might be, given its power in other hypnotic contexts – perhaps it is not as powerful as we thought, and there is a different factor at work? – but this heavily implies that hypnotic speech could benefit from some qualitative re-evaluation. Perhaps instant inductions should begin incorporating "relax" rather than "sleep" into their language. Alternatively, perhaps the word itself is not important, but rather it is the tonality and context that matters; there are many possible answers to this question, and it could benefit from further analysis.

It was also noted that mental imagery is often an important part of hypnotic induction, and as a result there is often mention of the hands in conjunction with other abstract senses, perceptions, feelings or thoughts, intended to focus the subject away from reality and onto something less tangible. Our results show that the words "feel," "mind," "time," (an extremely abstract concept) and "hand" are all key words for hypnotic language. A perusal of the scripts used for training shows that each of these words is used to direct attention toward similar abstract concepts, and "like" is often used to elicit similes or metaphors; this is imperative to evoking detailed mental imagery for a subject to focus on.

Let us now turn to an analysis of two of "that's" and "right". Post-analysis search through the corpus reveals that "that's" and "right" are used in just the way that you would expect them to be: "that's right". This appears to be an important part of hypnotic language that is often mentioned only briefly: encourage the subject. This phrase is particularly salient in those scripts which are from research sources ( Stanford Hypnotic Susceptibility Scale, Form C; Weitzenhoffer and Hilgard, 1962, Arizona Motor Scale of Hypnotizability; Kihlstrom, Glisky and McGovern), where safety and procedure are of much greater concerns, but the statistical prevalence of this phrase implies that it may constitute further analysis and consideration for increased use in hypnotic language.

So far, we have made very little mention of negation, and the use of "not" in hypnosis scripts. Additionally, we claimed that words like "those" and "that" should be more common in hypnotic scripts due to their dissociative properties, and their helpfulness in creating "out of body" experiences. Further analysis of statistical results produced by the model gave us the following counts for occurences of each linguistic variable:

| word | No. in hypnotic | No. in not hypnotic |
|---|---|---|
| not | 78 | 112 |
| that | 420 | 235 |
| those | 37 | 16 |

**Figure 3:** Appearances of negation ("not") and dissociative terms ("that," "those")

The statistics in Figure 3 show that the theoretical claims concerning dissociative and negative terms are consistent with the structure of hypnotic scripts: "those" and "that" appear about twice as much in hypnotic scripts, and although "not" is often a necessary linguistic variable in expressing many basic concepts, there does appear to be a conscious effort to reduce its use in hypnotic language.

## Conclusion

The computational analysis of hypnotic scripts presented in this study has served to support some of the previously discussed theoretical claims about hypnotic language, as well as to qualify others. A possible direction for this research could be to develop hypnotic scripts that are more consistent with these computational findings in an effort to improve upon our knowledge of hypnotic language. Additionally, further computational analysis of hypnotic scripts could provide a better understanding of the cognitive mechanisms involved in states of trance.

There are several possible steps in improving this analysis. Firstly, comparison of hypnotic documents to other types of documents, such as instructional documents, could yield interesting results, as it may be more difficult to distinguish between a list of instructions and a hypnotic induction, and further linguistic intricacies may be discovered. Furthermore, an examination of word order and sentence structure may also shed light on the distinctions between hypnotic language and standard speech. Also, in this analysis, we did not stem words (i.e. we did not transform "relaxed" to "relax"; note that these are both common words in hypnotic scripts), and an implementation of a similar model which ignores derivational or inflectional differences could provide stronger statistical information.

## References


Brahaj, A. (2009). *List of english stop words.* http://armandbrahaj.blog.al/2009/04/14/list-of-english-stop-words/.

Gafner, G., & Benson, S. (2000). *Handbook of hypnotic inductions.* New York, London: W.W. Norton & Company.

Hammond, D. (1990). *Handbook of hypnotic suggestions and metaphors.* New York: W.W. Norton.

Hewitt, W. W. (2003). *Hypnosis for beginners.* St. Paul, Minnesota: Llewellyn Publications.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics. 2nd edition.* Prentice-Hall.

Kihlstrom, J. F. (2007). Consciousness in hypnosis. In M. M. Philip David Zelazo & E. Thompson (Eds.), *The cambridge handbook of consciousness.* Cambridge.

Kihlstrom, J. F. (2008). The domain of hypnosis, revisited. In M. Nash & A. Barnier (Eds.), *The oxford handbook of hypnosis.* Oxford University Press.

Kihlstrom, J. F., Glisky, M. L., & McGovern, S. (1994). *Arizona motor scale of hypnotizability.* University of Arizona.

Weitzenhoffer, A., & Hilgard, E. (1962). *Stanford hypnotic susceptibility scale, form c.* Consulting Psychologists Press.