

springer@daf.lmu.de

Estrangement and Literary Information Content

Dr. Matthias Springer



Computing the average information content of a message:

$$H = - \sum p_i \log_2 p_i$$

- To suggest an idea for measuring features like the literariness or narrativity of a text
- Conceptualization of a theory of measurement
- This theory is borrowed from the mathematical aspects of information theory
- In order to find the information content of a message, the rule is:
 1. Translate the message in the binary system of a computer
 2. Count the number of zeros and ones

How to proceed?

- To conceptualize a textual feature which points to literariness
- To operationalize it in terms of the binary system → structuralism works with it
- Analyse and mark the features within the text
- Tag the feature when it appears marked as literary
- Compute the likelihood of the textual features to appear unmarked or marked respectively
- Fill in Shannon's formula

- Content Points to quantity (**Gehalt**), not to quality (**Inhalt**)
- Asks for an amount, not for semantics
- A certain amount of text is considered as literary significant in a strong quantitative sense
- Theory of literariness fits with this aspects

- Theory of literariness
- Simone Winko: Three Concepts
 - Based on linguistic features → V. Sklovskij, R. Jakobson
 - Terms & Conditions which qualifies a text as literary
 - Reader oriented concepts
- My Focus: Slovskij and Jakobson
Literature uses a special kind of language which differs from our every day language which means while you read:
 - Estrangement
 - Defamiliarization
 - Deautomatization

- Count the words of a text as the number of tokens
- Marking tokens with several characteristics
 - Unexpected
 - Stylistic figures
 - Interjecting the progress of „streaming“ through the text → deautomatization
 - Kind of estrangement
- Rule: Do not think about the effect and semantics of each marked feature, just detect it
- Count the number of marked or unmarked tokens respectively

Strong quantitative
empirical text features
on the level of language
level as an expression of
literariness

L
i
t
e
r
a
r
y

I
n
f
o
r
m
a
t
i
o
n

Probability of
appearance of a symbol
as the data basis to
calculate the information
content

The significant aspect is that the actual message is one selected from a set of possible messages.

Indirect relationship:

removed average information content =
obtained information



- Uncertainty: expression by estrangement
- Uncertainty raises up literariness
- Uncertainty : Expression through probability
- Uncertainty raises entropy

Requirement: Transforming data of literariness into data of information theory

- Each token is one stochastic event
- Each event could be successful or fail:
 - Fail: no estrangement → automatized reading
 - Success: estrangement → deautomatized reading

- Binominal Distribution
- Abstract and simple model
- Independent events
- Condition: same likelihood for each event
- Order of successful events is ignored in this distribution
- B.D.: First approach to a rule of measurement
- Calculation of Probabilities according to the number of successful events

Es sind im Gebirge ungeheure Materialströme vorhanden, welche nur der ordnenden Hand eines Helfers, der, leidenschaftslos, sich niemals von Vorlieben oder Abneigungen leiten ließe, bedürfen, um ihr Geröll, Geschiebe, Erdreich, und ihren Schlamm loszulassen, diese Leckereien, die das Gebirg eigentlich ganz allein essen wollte. Es hat uns nichts davon abgeben wollen. Doch jetzt überreicht uns die Bergwelt mit einer leichten Verbeugung sich selbst sowie die ganze Umgebung dazu. So hat der Urlauber es sich nicht vorgestellt. Daß er, lüstern, begeistert, einem Verhältnis zur Natur auf der Spur, nicht sie, diese Umgebung, betrachtet, sondern daß vielmehr sie auf ihn fliegt.

Der brave rundliche Ausflugshügel mit den Urwaldmassen drauf läßt diese jetzt von sich abgleiten, mit einer Bewegung, nachlässiger, als wenn man eine Fliege wegwedeln würde. Der Förster trifft noch eine Anstalt, sich zu retten, aber er kann dann nicht mehr hineingehen. Die weißen Zipferln seiner Söhne (wie klein waren die einmal! Man erinnert sich kaum) haben für den Bruchteil einer Sekunde noch, zwei Kerzen, zwei Grableuchten, Licht gespendet, oder warens nur zwei helle Steine am Rand des Abgrunds? Der Wald, erzwungen von der elementaren Macht der Suche nach einem Ausweg, marschiert jetzt ins Tal hinunter, er kommt, er kommt! und es kommt mit ihm durchs nagelneue Bachbett, wo das alte Geschiebe und die alten Moränenreste besonders labil geschichtet sind, weil sie ja schon mehrfach, jedoch zumindest einmal, aufgrund der Unwetter der letzten Wochen, beim Schlafen gestört worden sind, sich den Wasser- und Erdzutritt erzwingend: DIE MURE. DIE FURIE.

Die Erde gibt sofort den Widerstand gegen das Wasser auf und geht gleich mit ihm mit. Der Regen ist wie ein Mensch, der an Gott glaubt, er weiß, er muss sich keine Sorgen machen, er wir schon irgendwo ankommen.

Statistik:

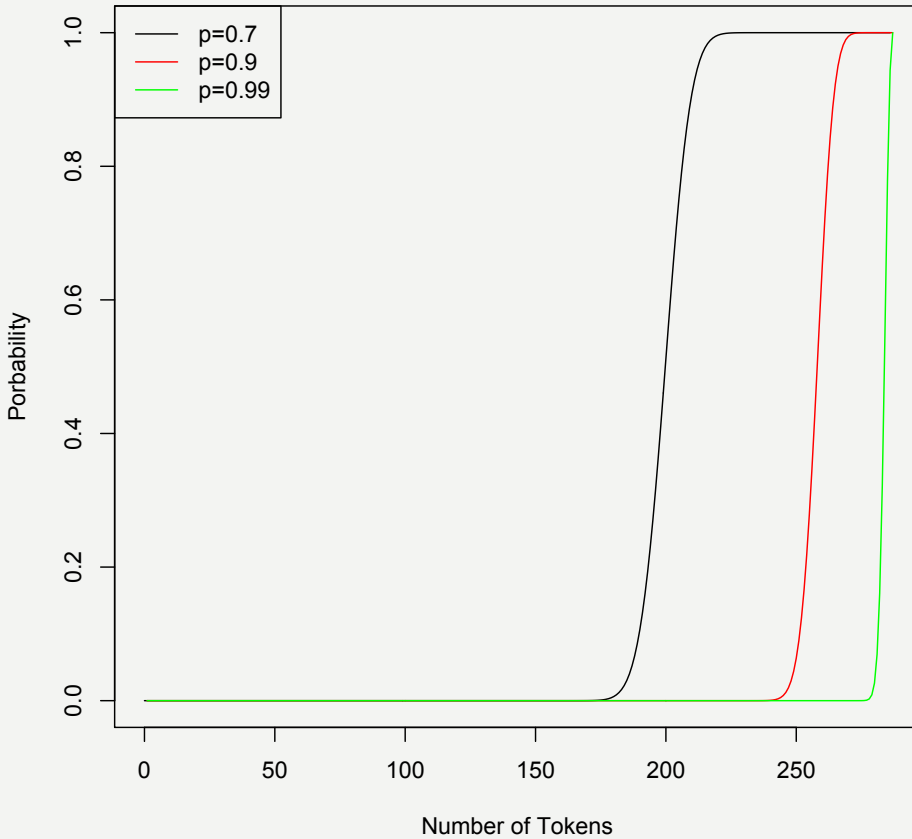
286 Wörter

76 Tokens unmarkiert, ließen sich also als automatisiert in der Leserwahrnehmung bestimmen.

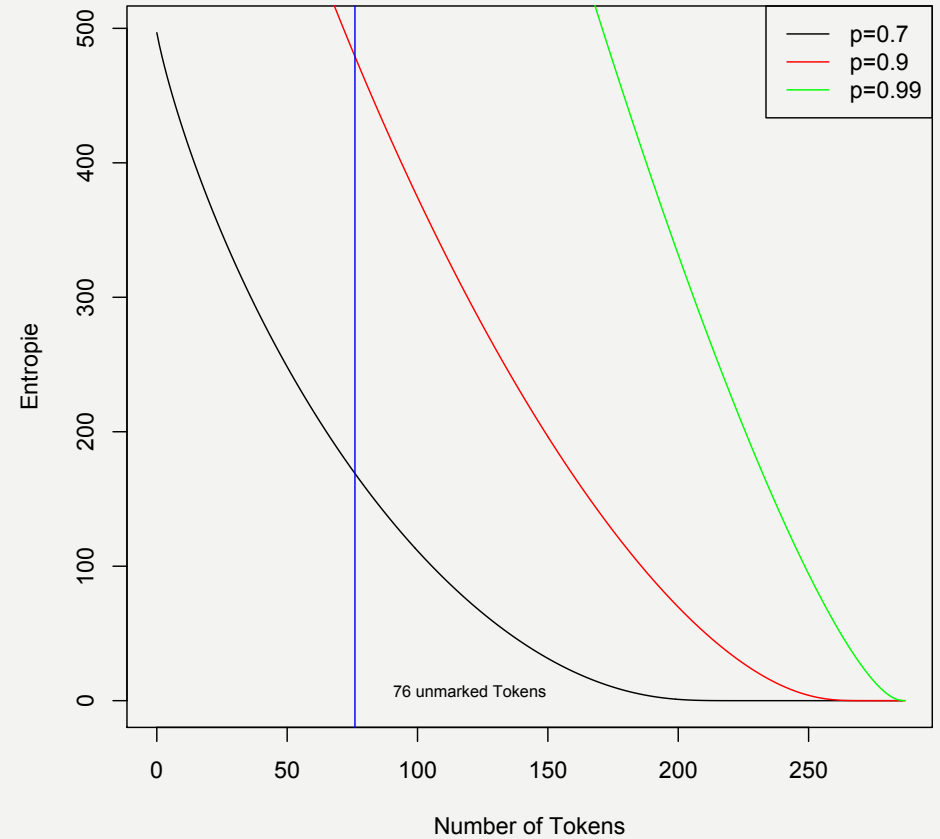
- 286 tokens
- 76 tokens unmarked / 210 tokens marked

286/76	P=0.7	P=0.9	P=0.99
Prop	1.155499e-51	1.453068e-143	0.000000e+00
EntropieProp	169,21	474,50	Undef.
cumProp	1.365628e-51	1.513599e-143	0.000000e+00
EntropieCum	168,97	474,44	Undef.

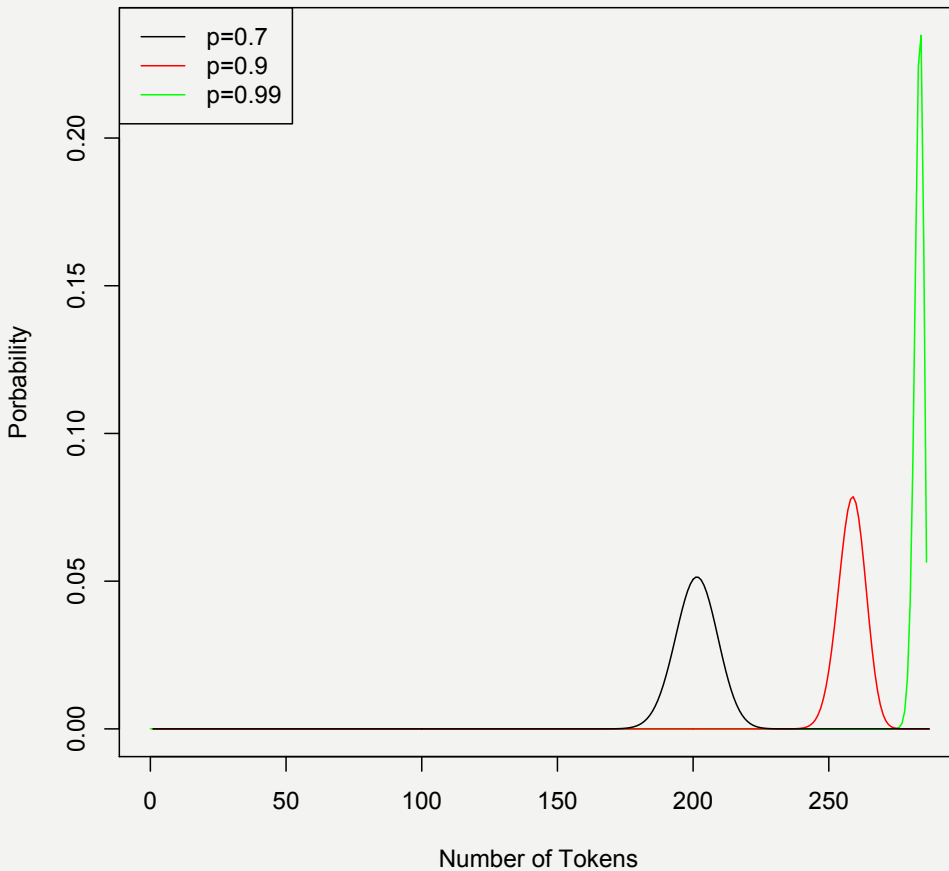
Cumulative Probability Distribution



Entropie

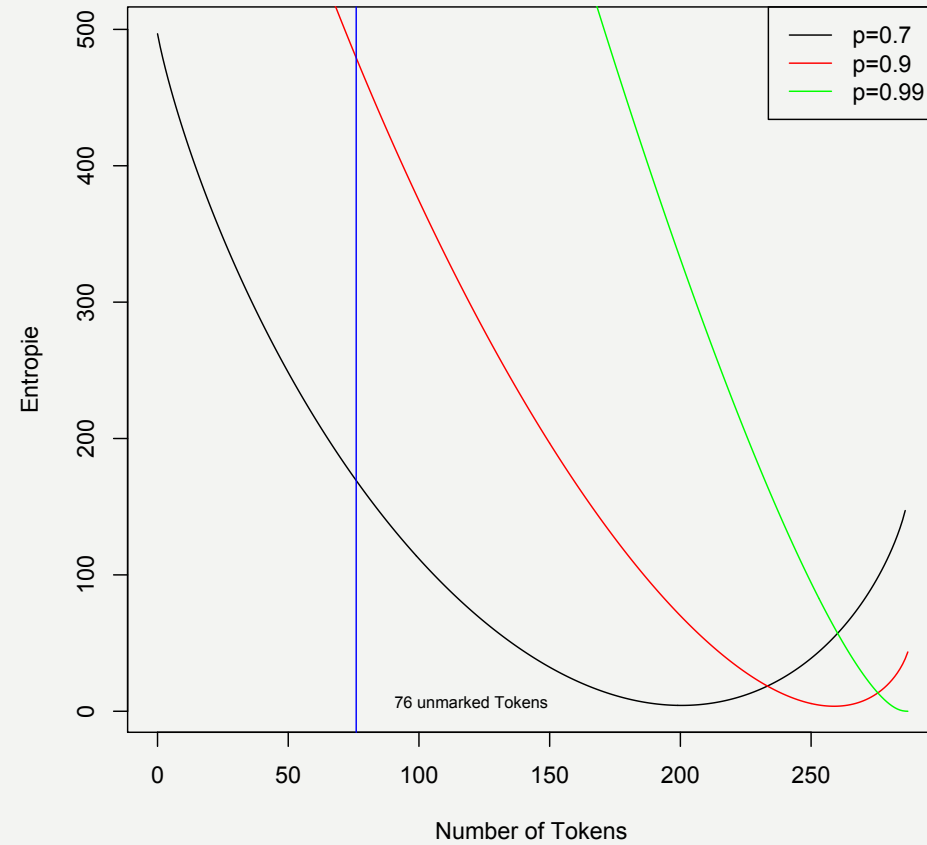


Probability Distribution



Binomial Distribution of unmarked/marked Tokens with $N=286$

Entropie



Mein Freund A. erlitt eine Reihe von Niederlagen in der Liebe. Aber je öfter er scheiterte, desto komplizierter wurde die nächste Frau, die er suchte. Nach einer Lesung trafen wir uns auf ein Glas Wein. Er hatte sich gerade von seiner treuen, schönen, klugen, sportlichen, häuslichen und **natürlichen** Frau getrennt, weil sie **schweigsamer** gewesen war **als ein Grab**. Nun beschrieb er mir seine Traumfrau, die er, endlich gereift, verdient **habe**. Eine Mischung aus Mutter Theresa, Marie Curie, Clara Schumann, Marilyn Monroe, aus Talkmasterin, Ballettänzerin und einer **modebewußten Untergrundkämpferin**. Ich mußte lachen und wunderte mich nicht, daß viele Frauen das Weite suchten, wenn ihnen die Hirngespinnste meines Freundes klar wurden. Doch plötzlich meldet sich A. bei mir. **Überglücklich**. Er **hätte** es selbst nicht geglaubt, aber die Frau seiner Träume **sei** da. Sie **sei** seine Freundin. **Unglaublich!** Zum ersten Mal **erlebe** er, wie sogar die **trockenen** Worte Hegels **erotisch wirkten**, wenn sie sie **zitieren**.

Einen Monat später allerdings **weinte** mir A. den **Telefonhörer voll**. Nein, **nein**, die Frau **sei** wirklich ein Traum, **doch habe** er nur noch schlaflose Nächte und **frage** sich, warum eine solche Frau ausgerechnet so einen **miesen** Typen wie ihn **liebe**, wo doch alle Welt ihr zu Füßen **liege**.

Statistik:

200 Wörter

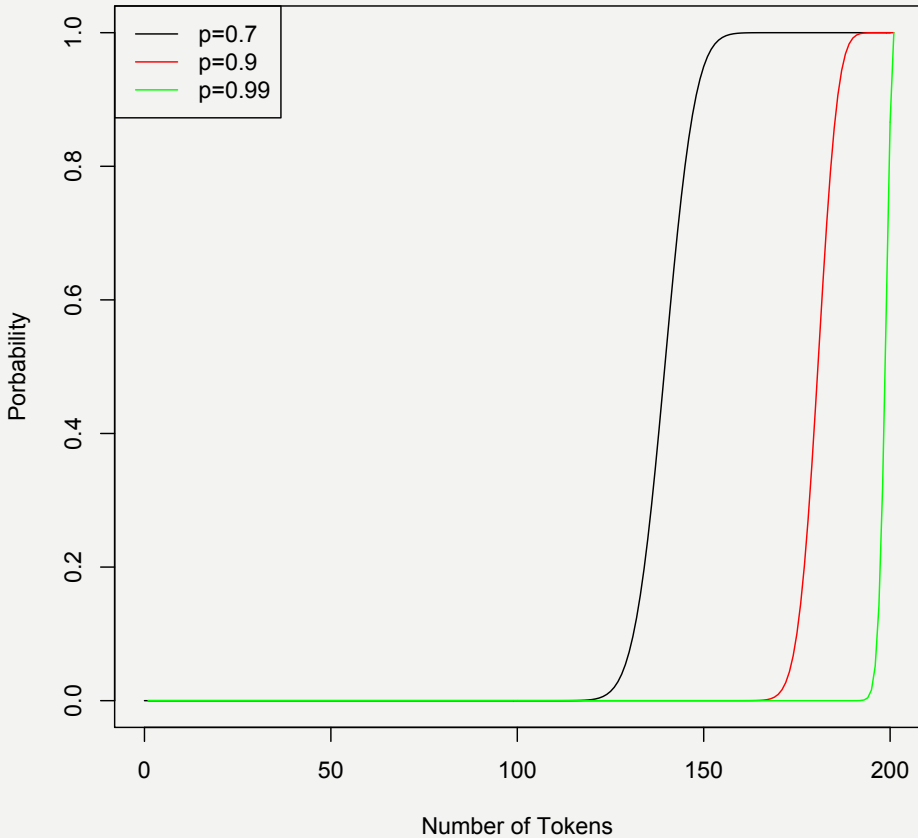
29 Tokens markiert, ließen sich also als ent-automatisiert in der Leserwahrnehmung bestimmen.

171 Tokens sind unmarkiert

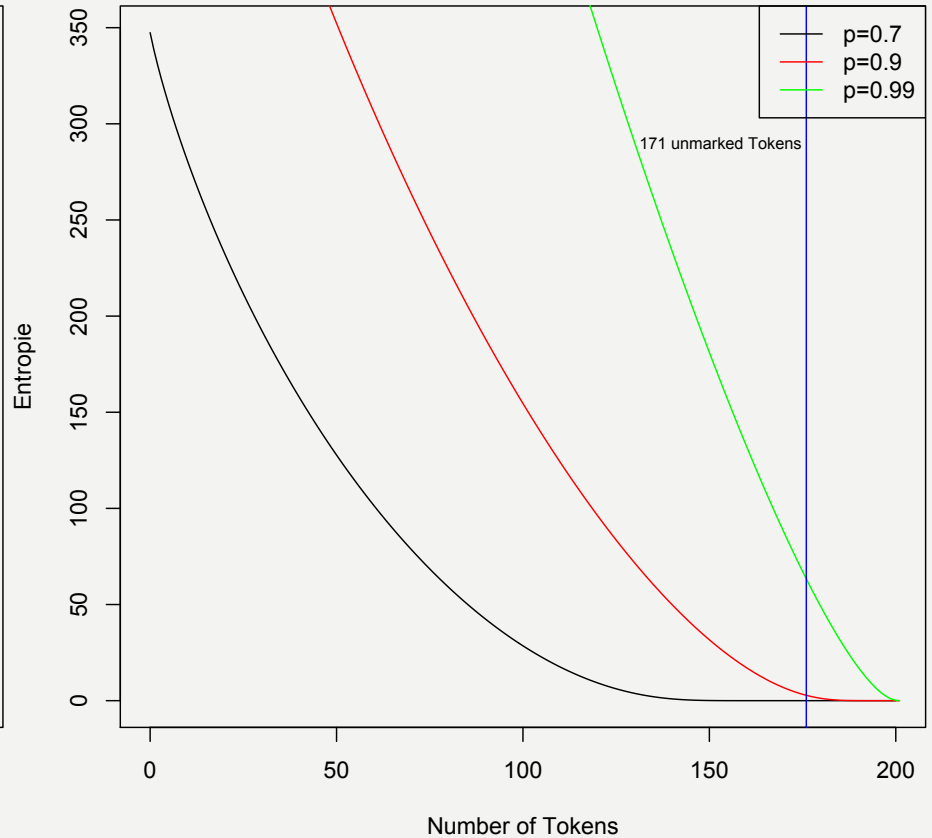
- 200 tokens
- 171 tokens unmarked / 29 tokens marked

200/171	P=0.7	P=0.9	P=0-99
Prop	1.602703e-07	1.076566e-02	1.288813e-24
EntropieProp	22,57	6.54	79,36
cumProp	9.999999e-01	2.709223e-02	1.367359e-24
EntropieCum	1.441386e-07	5,21	79,27

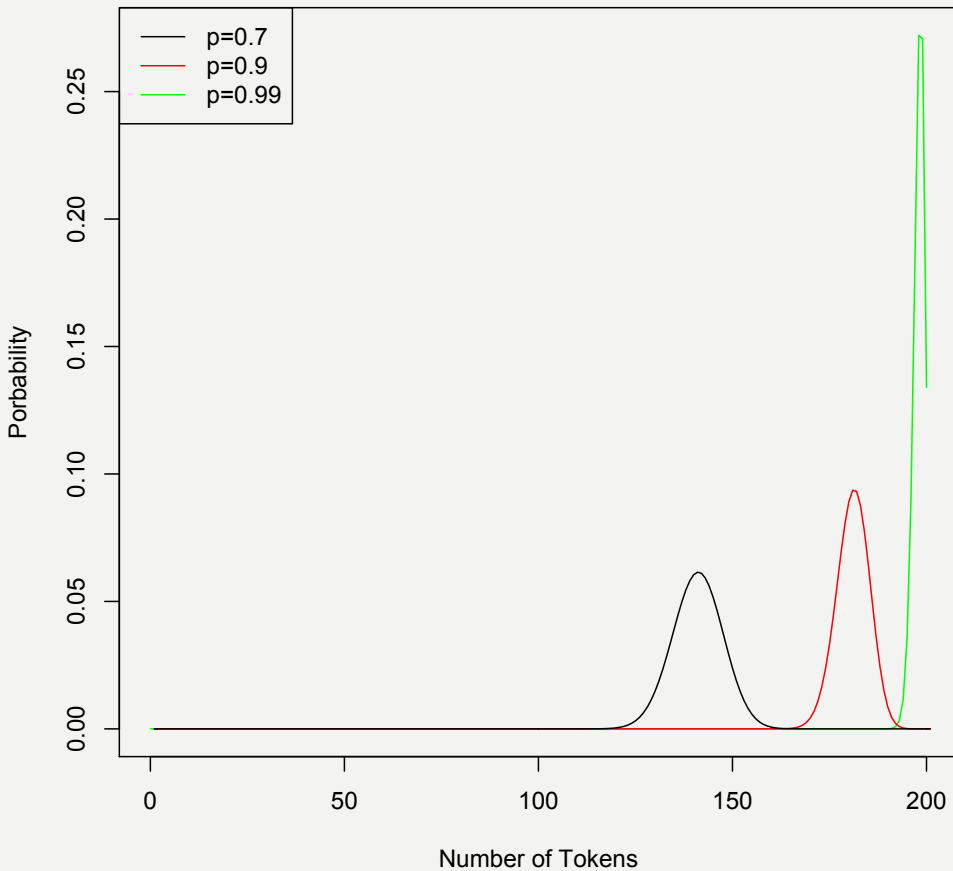
Cumulative Probability Distribution



Entropie

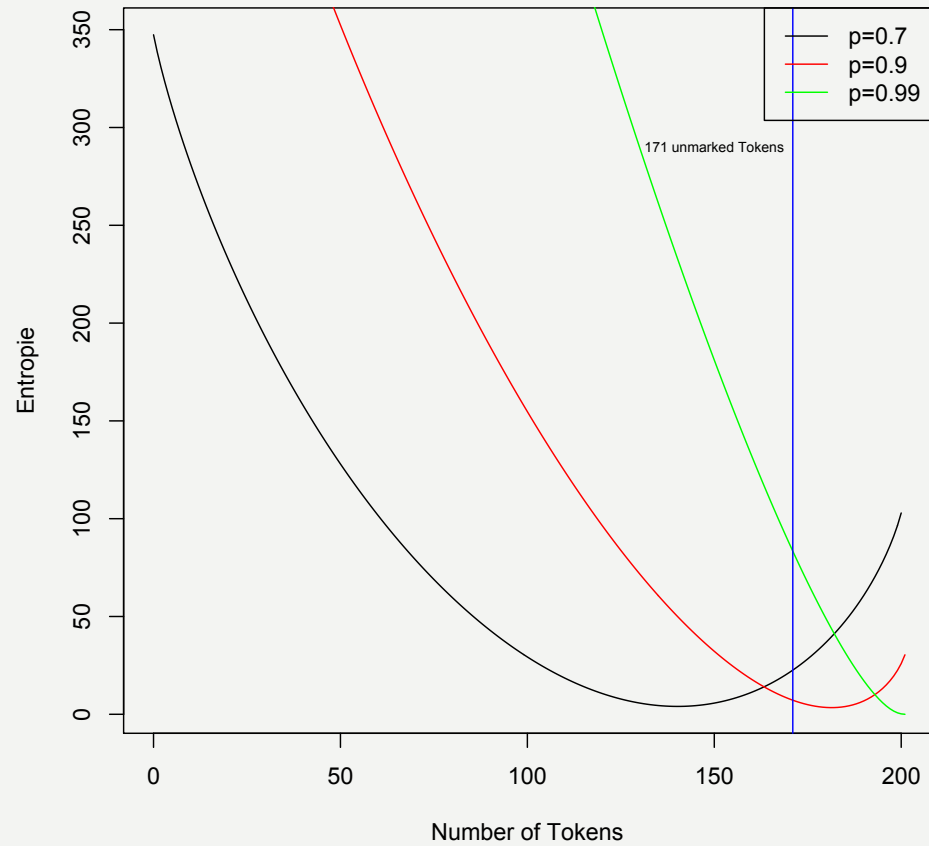


Probability Distribution



Binomial Distribution of unmarked/marked Tokens with N=200

Entropie



- Binomial distribution is not the correct model:
 - Different likelihoods must be assumed for single tokens when detected estranged
 - likelihood is determined arbitrarily by myself and not by convention or empirical observation
 - Tokens are not independent, they are highly dependent to their context
- Concept of estrangement is not really operationalized by myself
 - Further theoretical work is needed
 - Parsing the text needs to be done by empirical readers
- A study for validation needs to be executed

- ... I Think with my general assumption about
- the relationship between literariness and information theory
 - I get somewhat I will call literary information.
 - This concept offers us a way to measure literary quantity
 - In order to make Predictions about the literary quality
 - and a potential literary response by the recipient of a text.

Thank you for your interest