

springer@daf.lmu.de

Average Information Content

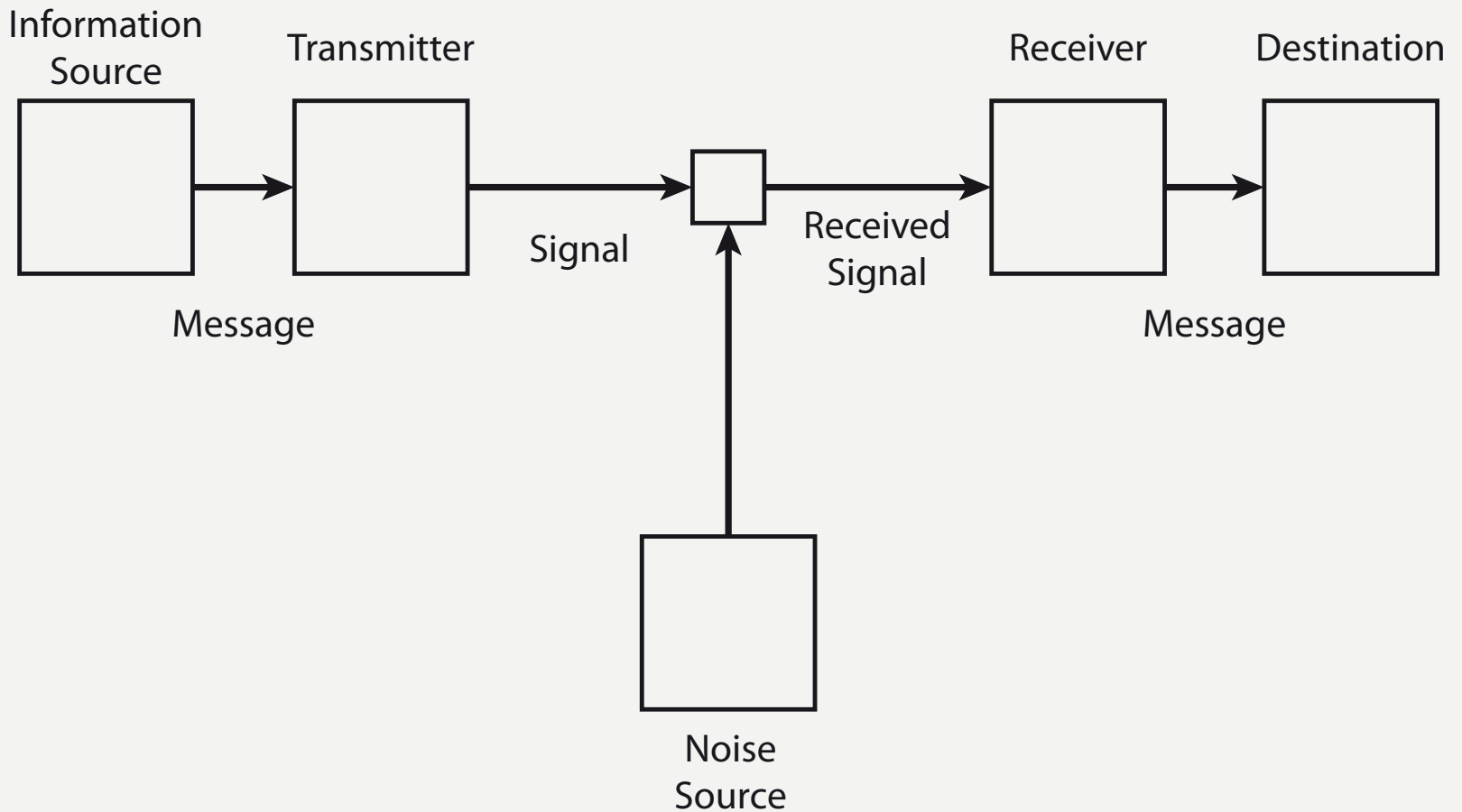
Matthias Springer

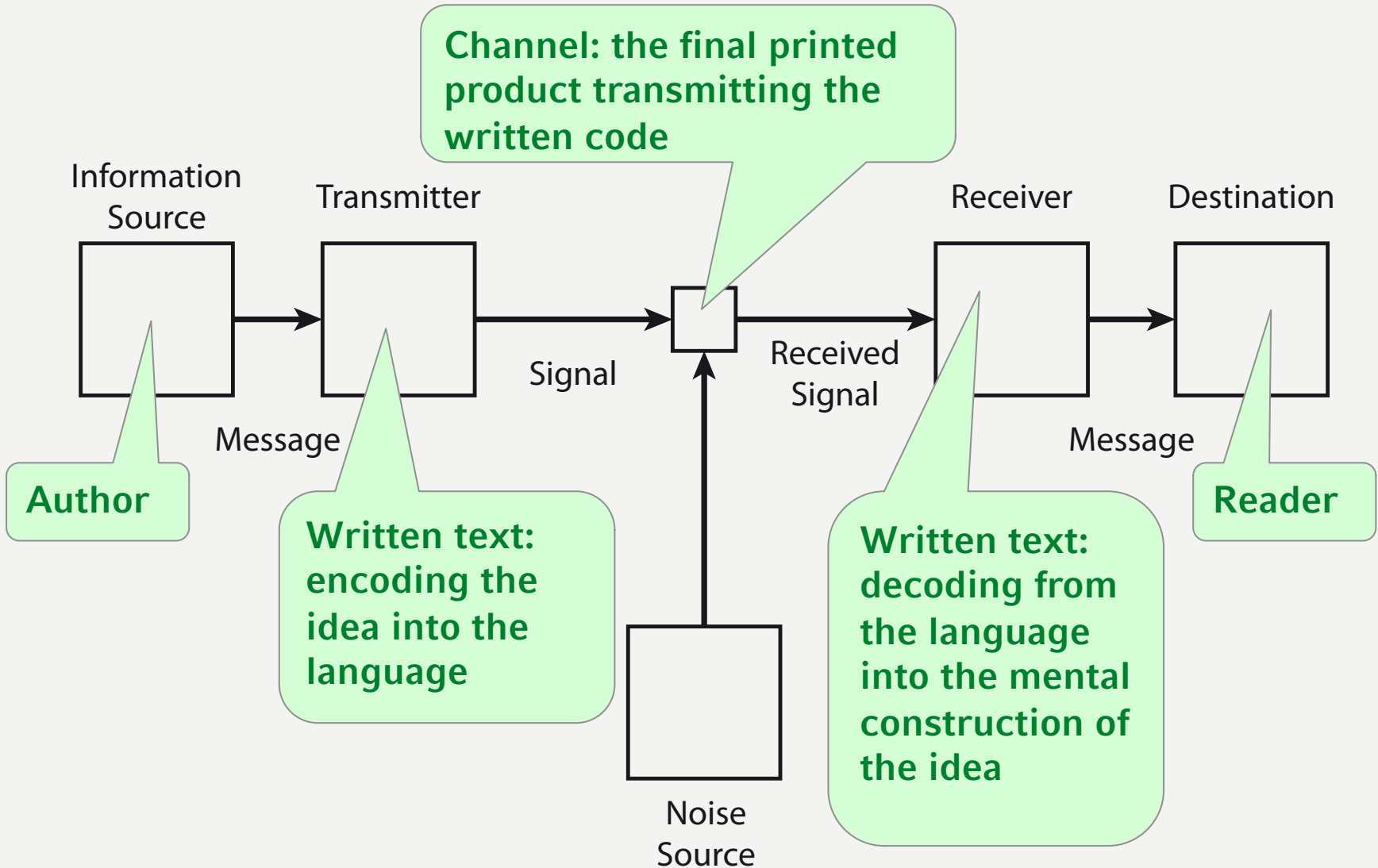


- Two parts
- Introduction to average information content as a theory of measurement
- A study showing how this could be applied to literature

- 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

- Each literary or poetical text feature is seen as a piece of literary information coded within the text
- A basic approach to this idea was provided by S. J. Schmidt who conceived literature as a subsystem of society based on communication – aesthetic communication
- Hence literature becomes a message and the literary process from the author to the reader becomes a process of communication
- Semantic aspects of a text are not relevant for measuring information content with a mathematical concept







“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. *These semantic aspects of communication are irrelevant* to the engineering problem. *The significant aspect is that the actual message is one selected from a set of possible messages.* The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.”

(Shannon, 1948)

- Level A: How accurately can the symbols of communication be transmitted? (technical / statistical problem)
- Level B: How precisely do the transmitted symbols convey the desired meaning? (semantic problem)
- Level C: How effectively does the received meaning affect conduct in the desired way? (effectiveness, pragmatic problem)

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

- The interpretation of a literary feature is not relevant for its literariness but how it is encoded as a symbol and what its likelihood to appear is (Level A)
- This means the question whether the symbol enables the intended effect in the receiver potentially (Level C)

Computing the average information content of a message:

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

- Flip a coin: you will get either heads or tails
- For each event which appears you can count one *bit* of information
- If you flip the coin twice, you count two bits of information by four possibilities; with three flips three bits by eight possibilities
- In general: the coin-system contains x bits of information by y flips (remember C. Shannon)
- This is written by the function: $y = 2^x$
- To compute the average information content the inverse is used: $x = \log_2 y$

- The mathematic relation within the formula:

$$\log_2 \frac{a}{b} = -\log_2 \frac{b}{a}$$

- Likelihood expresses the potential of an actual information occurring within a set of possible messages that it will appear based on its probability
- Likelihood considers more complex situation e.g if the symbols in a system are not distributed uniformly:
 - a) Role a dice twice and write the serial of numbers: each possible serial comes up with the same likelihood of 1/36; this message contains 5,17 bit of average information content
 - b) Role a dice twice and count the numbers: the event "2" appears with a likelihood of 1/36 and "7" with 1/6; this message contains 3,27 bit of average information content
- System a) contains more average information content than system b), or contrarily system b) contains more obtained information than system a)

Indirect relationship:

removed average information content =
obtained information

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

Matthias Springer

Humor aus dem Computer?

Grundlagen einer Web-Applikation
zum Test von Komik in narrativen Texten

Basic unit: Event

***Change of the state of an object in the
chronological progress of the action***

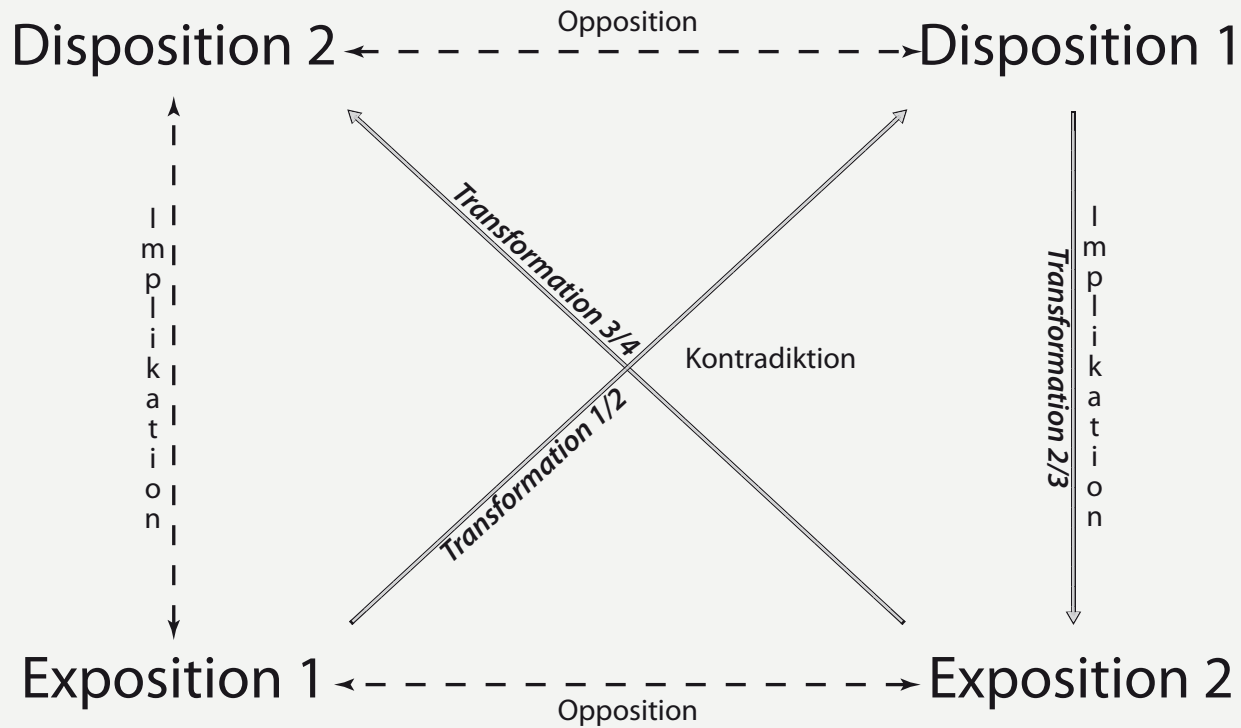
Basic complete action: Episode

Built out of two integrated events

Event(Object, P_Exp, Pclass_Exp, E_Start,
E_End, P_Dispatch, Pclass_Dispatch, D_Start, D_End)

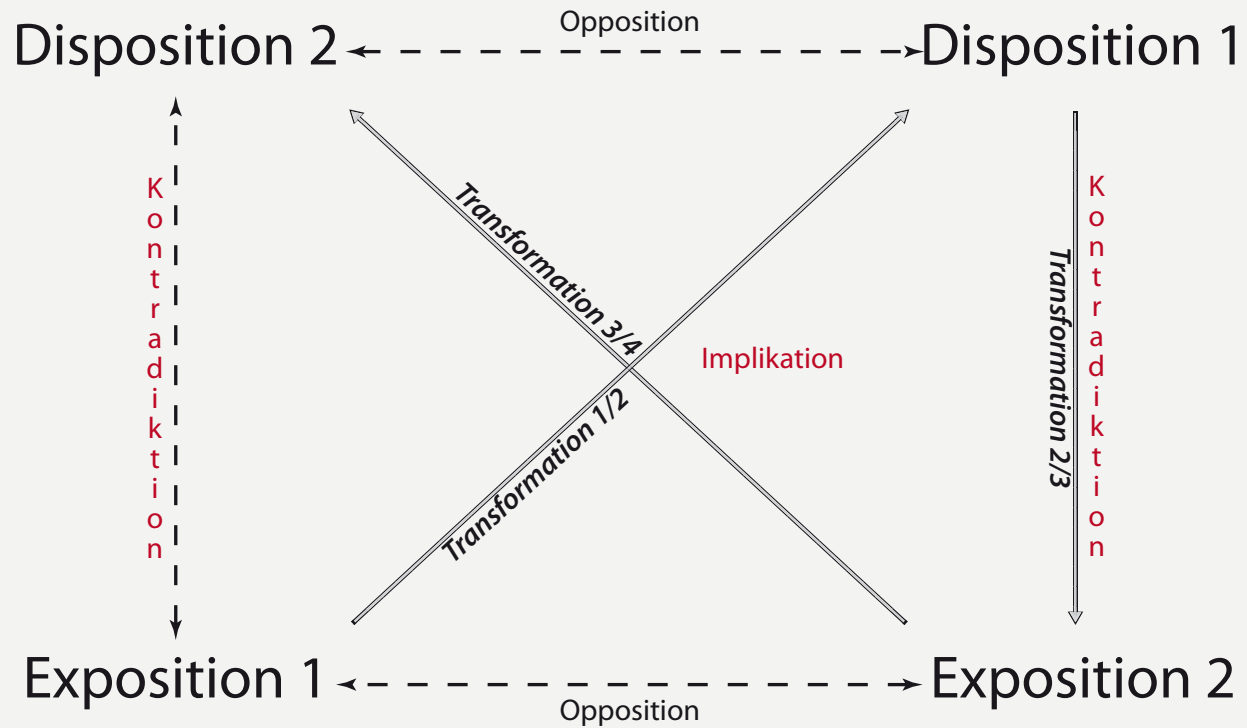


Adaption von Algirdas Greimas semantischem Quadrat auf eine Episode nach Meister (2003)





Das semantische Quadrat einer potentiell komischen Episode



1. The material:

- Hebbel: Unverhofftes Wiedersehen (prediction: no pot. humor action)
- Hebbel: Kannitverstan (prediction: pot. humor action)
- Boccaccio. Novel of Peronella (from Decameron, prediction :pot. humor action)

2. Parsing events by trained subjects:

- Identification of an object
- Marking the expositional phrase
- Marking the dispositional phrase

3. Coding by other subjects

- Expositional phrase with an semantic term
- Dispositional phrase with an semantic term
- Defining the the logical relation between exp. and disp. as contradiction (“normal”) or implication („pot. humorous“)

4. PostHoc questionnaire:

- Evaluation of each narration with adjectives
- Evaluation of the reading experience with validated and reliable scales (Groeben e. al.)

5. Analysis

- Test of the hypothesis, that the increasing of average information content of action and the evaluation of the narration as humorous correlates.
- Modelling the size of this relation

Promotion Matthias Springer

anzuzeigen, warum diese Personen nicht möchten ehe-lich zusammenkommen", da meldete sich der Tod. Denn als der Jüngling den andern Morgen in seiner schwarzen Bergmannskleidung an ihrem Haus vorbei-ging, der Bergmann hat sein Totenkleid immer an, da klopfte er zwar noch einmal an ihrem Fenster und sagte ihr guten Morgen, aber keinen guten Abend mehr. Er kam nimmer aus dem Bergwerk zurück, und sie saumte vergeblich selbigen Morgen ein schwarzes Halstuch mit rotem Band

Arbeitsbereich

Ereignisse noch auszuzeichnen: 8 von 8

Ereignisse interpretieren:

Das Interpretieren der Ereignisphrasen (Exposition und Disposition) erfolgt in 4 Schritten:

1. Auswählen einer Prädikatklasse
2. Auswählen oder Eingeben eines Prädikat-Terms
3. Festlegung der Beziehung der Prädikate von Exposition und Disposition
4. Speichern der Interpretation (diese kann durch Auswählen des Ereignisses im Ansichtsbereich jederzeit geändert werden)

Ereignisobjekt:	Prädikatsklasse:	Prädikat:
<p>Eheversprechen</p> <p>Exposition: In Falun in Schweden küßte vor guten fünfzig Jahren und mehr ein junger Bergmann seine junge, hübsche Braut und sagte zu ihr: »Auf Sankt Luciä wird unsere Liebe von des Priesters Hand gesegnet. Dann sind wir Mann und Weib</p> <p>Disposition: sagte die schöne Braut mit holdem Lächeln, »denn du bist meineinziges und alles, und ohne dich möchte ich lieber im Grab sein als an einem andern Ort.</p>	<p>sozial (beherrscht / beherrschend)</p> <p>Klasse wählen</p>	<p>Neuer Prädikatsterm abgelehnt angeboten angenommen</p> <p>Ändern</p> <p>Ändern</p>
<p>53 In welcher Beziehung stehen die Prädikate der Exposition und Disposition?</p>	<p><input type="radio"/> Gegensatz (Tag - dunkel)</p> <p><input checked="" type="radio"/> Entsprechung (Tag - hell)</p>	

Auszeichnung speichern

Ansichtsbereich

Ereignisse anzeigen: alle Ereignisse

Event-ID	Objekt	P-Klasse-Exposition	Prädikat-Exposition	P-Klasse-Disposition	Prädikat-Disposition	P-Beziehung
----------	--------	---------------------	---------------------	----------------------	----------------------	-------------



2

Verwunderung noch größer, und wenn ein Berliner sich über den Verkehr beschwert, denken die Brasilianer, er mache einen Scherz. In Brasilien ist alles oft genau das Gegenteil von dem, was in Deutschland geschieht. Es heißt, wenn zwei Deutsche eine Straße ohne Ampel überqueren und einer von ihnen sich erschrickt, weil plötzlich ein Wagen auftaucht, dann sagt der andere: „Nicht erschrecken, der hat uns schon gesehen!“
 Man erzählt sich die Geschichte eines Ausländers (wer weiß, vielleicht eines Deutschen) in einem Taxi in São Paulo, der dauernd die Augen schloss, weil der Fahrer an keiner roten Ampel hielt. Aber bei der ersten Ampel, die auf Grün schaltete, hielt der Fahrer einen Augenblick. Entsetzt fragte der Gast, warum. „Ach, bei Grün muss man halten“, erklärte der Fahrer, „weil manchmal ein Verrückter aus der anderen Straße kommt.“
 Man könnte also meinen, eher ertrinkt ein Olympiaschwimmer in einem Kinderschwimmbassin, als dass ein Brasilianer in Berlin angefahren wird. Weit gefehlt. Wir haben beide, meine Frau und ich, noch immer das Überqueren der am stärksten

Ich fand den Text ...

3

01: unterhaltsam	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	gar nicht	wenig	mittelmäßig	überwiegend	völlig
02: humorvoll	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	gar nicht	wenig	mittelmäßig	überwiegend	völlig
12: bedrückend	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
	gar nicht	wenig	mittelmäßig	überwiegend	völlig

Bitte geben Sie noch an, wie es Ihnen während des Lesens ergangen ist.

4

13: Die Leseerfahrung war für mich intensiv.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	gar nicht	wenig	mittelmäßig	überwiegend	völlig
14: Gut, dass der Text nicht noch länger war.	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
	gar nicht	wenig	mittelmäßig	überwiegend	völlig

- To compute the average information content to calculate the informational entropy out of the marked episodes
- Adding the average information content of ambiguous episodes computed out of events which are marked as potentially humorous (relation of implication)

	E_n	E_k	$E_n \sim E_n$	$E_k \sim E_k$	$E_n \sim E_k$
N_E	6	3	$\binom{6}{2} = 15$	$\binom{3}{2} = 3$	$6 * 3 = 18$
$p_{Episode}$			$15/36$	$3/36$	$18/36$

$$H = -\log_2\left(\frac{1}{36}\right) = 5.17 \text{ bit}$$

General average information content

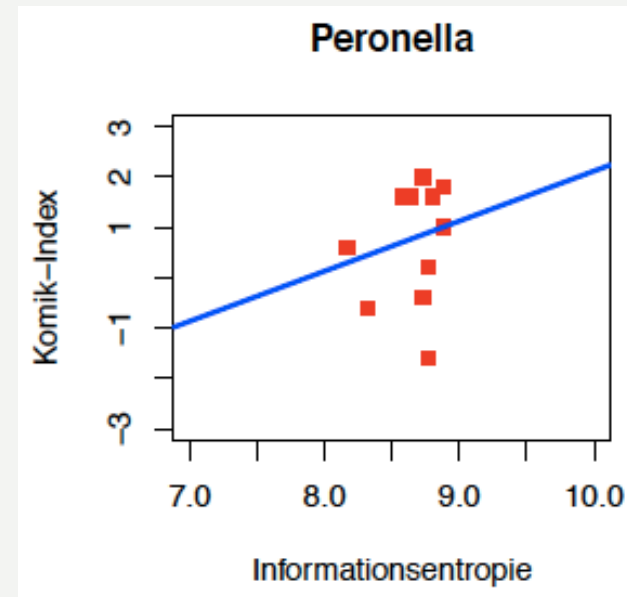
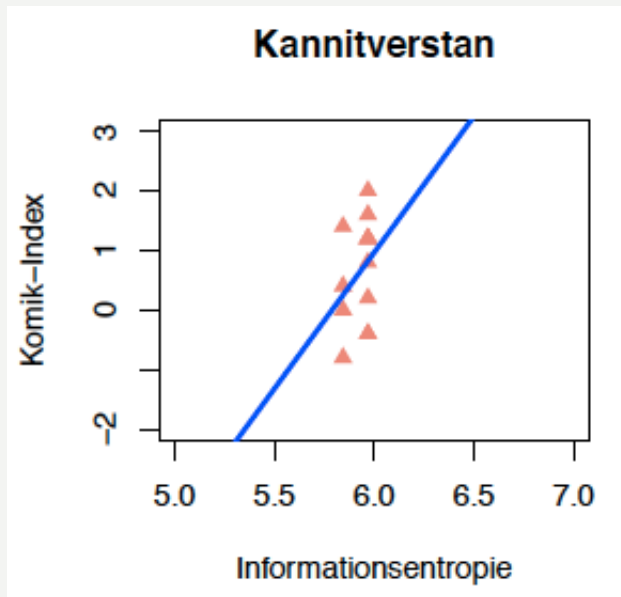
$$H_a = -\frac{3}{36}\log_2\left(\frac{3}{36}\right) - \frac{18}{36}\log_2\left(\frac{18}{36}\right) = 0.8 \text{ bit.}$$

Extended average information content

$$H = H_e + H_a = -\log_2\left(\frac{1}{36}\right) - \frac{3}{36}\log_2\left(\frac{3}{36}\right) - \frac{18}{36}\log_2\left(\frac{18}{36}\right) = 5.97 \text{ bit.}$$

Total average information content

A relation between the average information content and humor feeling is indicated



- The concept of noise?
- Some shortcomings in my study:
 - Low number of subjects
 - Parsing the events
 - Coding and Tagging the events
 - Re-constructing the story: linking the the events by semantics
- In general: Conceptualization and operationalization of other textual features which are indicators for aesthetics

Thank you