Extraction and Analysis of non-canonical Questions from a Twitter-Corpus

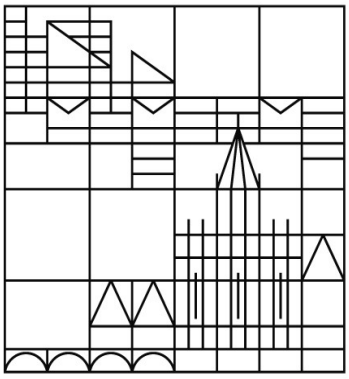
Master-thesis

presented by:

Zymla, Mark-Matthias

at the

Universität Konstanz



Department of Linguistics

1. Evaluated by Prof. Dr. phil. Miriam Butt
2. Evaluated by Prof. Dr. phil. Maribel Romero

Constance, 2014

Special thanks to:
Isabel Riedel
Andreas Walker
Daniela Wochner

Content

I. Introduction ….....2

II. Theoretical part 5

1. Literature overview 6

1.1. Primary definition and descriptive features 6

1.2. Tests for rhetorical questions 11

2. Semantics and pragmatics 15

2.1. Semantics of questions 15

2.2. Rhetorical questions as assertions 17

2.3. Rhetorical questions as proper questions 20

2.4. Rhetorical questions in the common ground 22

2.5. Relatives of rhetorical questions 27

2.6. Conclusion 30

3. Research - study at the long night of sciences 31

3.1. The survey - procedure and method 31

3.2. The survey - results and discussion 33

III. Practical part 37

1. Challenges 38

2. The script 41

2.1. Resources 41

2.2. Input 44

2.3. Preprocessing of the raw-data 45

2.4. Detection mechanisms and the rating system 50

2.5. Evaluation 54

3. Future possibilities 55

3.1. Implementation of sophisticated grammar systems -XLE 55

3.2. Implementation of filed semantics and pragmatic systems 57

3.3. Conclusion 61

IV. Literature 62

IV. Appendix 64

I. Introduction

This thesis deals with the phenomenon of rhetorical questions. This type of question can be easily differentiated from information-seeking questions, as can be seen in in (1). Rhetorical questions are most commonly described as a persuasive tool of communication in the shape of a question, but with the illocutionary force of an assertion. However, their functionality is still not fully understood or rather formalized.

(1)	a) Who wouldn't like to drive Ferrari?	RQ
	a') Everybody likes to drive a Ferrari.	Underlying assertion
	b) Where is the concert hall?	ISQ

In the course of this thesis, two main goals are pursued: On the one hand, mapping rhetorical questions on the landscape of Semantics and Pragmatics to describe both the assertive/declarative character of rhetorical questions and their persuasive aspect. On the other hand, exploring natural as well as machine language processing strategies to detect rhetorical questions in the environment of German social media. This will serve as preliminary work in the automatic analysis of rhetorical questions.

To achieve this goal, the thesis is divided into two parts, one theoretical and one practical. To begin with, this work will summarize the theoretical research on rhetorical questions. For this purpose, data and theoretical claims from the literature will be investigated and illustrated. As this thesis will focus on German social media, the focus of the investigation will be on rhetorical questions in German. However, cross-linguistic data from analyses of other languages, will also be considered. One of the most comprehensive studies in the field of rhetorical questions is provided by Jörg Meibauer. He offers an in-depth analysis of rhetorical questions in German. His work will be used as a starting point for the present analysis and will be especially helpful for the practical part of the investigation. Starting from Meibauer's work, a descriptive overview of rhetorical questions will be prepared. This will help to understand the strategies used to derive the meaning of rhetorical questions.

With this information the next task can be accomplished: describing the meaning of rhetorical questions. For this purpose, attention will be paid, firstly, to the question, which illocutionary force underlies rhetorical questions. In the present work, the position that rhetorical questions are indeed more related to *real* - information seeking - questions than to declaratives will be defended. This leads to the second point of focus: How can it be possible that although rhetorical questions are typically described as questions with the illocutionary force of a declarative or assertion, they can still be analyzed as questions. For this peculiarity a number of strategies will be proposed. Part of this will be the investigation of rhetorical questions in the framework of a common ground analysis, which will play an important role in two ways. The common ground helps to differentiate rhetorical questions from similar question types and it will also be used to describe another prevalent aspect of rhetorical questions, to which little attention is paid in current literature: their persuasiveness. For this purpose an approach using an extended common ground analysis along the lines of Gunglogson and Caponigro & Sprouse will be proposed.

The transition to the practical part will take place with some new data collected from a study prepared for this thesis. It is loosely based on a manually composed corpus of Tweets – short pieces of text of max. 140 characters length on the respective social media site Twitter - containing various types of questions. While the data can be examined from the theoretical point of view, it is more important for the practical part of this work. As a reminder: The main goal of the practical part is to develop and test automatic detection methods for rhetorical questions. These will be derived from the theoretical research as well as a study based on the work with Twitter, which provides information about detection mechanisms used in natural language processing.

In the analysis of the data gathered from the survey the influence of the origin of the data will become noticeable. This will make a description of the circumstances created by Twitter inevitable. This will also be the introduction to the second part of this work. Here the challenges and peculiarities of web-slang, especially in Twitter, will be laid out. Furthermore, the complex linguistic nature of rhetorical questions will be brought up again in the light of machine language processing.

For this thesis a script, written in Python, was developed to test some basic/shallow approaches to detect rhetorical questions in Tweets. The design and the functionality will be elaborated on, as well as the flaws and inevitable problems of these shallow approaches. The course of the analysis will lead to

the next step: developing more sophisticated systems to find rhetorical questions in tweets. This will only happen in the shape of thought experiments, since anything more would be going beyond the scope of this work. Ideas using deep grammar analysis, as well as proposals relying on semantic systems like ontologies will be depicted and their feasibility and their functional efficiency will be discussed. This overview over future possibilities will sum up the present work. In the appendix all collected data as well as a thoroughly commented version of the developed script will be offered.

II. Theoretical Part

1. Primary definition and descriptive features

This section will serve as a guiding point for the theoretical part of the present work. For this purpose two points are pursued. Firstly, this section will provide an overview over the descriptive literature on Rhetorical Questions(henceforth RQs) in German, beginning with an overtly distinctive type of RQ. This will help to set a proper reference point for the present thesis. Secondly, some cross-linguistic data from English is discussed. This information will help to roughly determine the functionality of rhetorical questions by describing them from various angles and serves as a cornerstone for the semantic and pragmatic analysis. It needs to be kept in mind that the first part of this chapter in particular is based on Rhetorical Questions in German. However, the language dependency will fade towards the semantic and pragmatic analysis encountered later in this work.

1.1. Primary definition and descriptive features

To begin with RQs are understood as Questions that have the illocutionary force of an assertion. This most prevalent aspect of RQs is also the most unanimously accepted fact in the research field. This is already shown in the comprehensive work of Jörg Meibauer who described both, the semantic and pragmatic aspects, based on speech-act-theory and the descriptive features of RQs. (Meibauer. 1986. Rhetorische Fragen. Tübingen. Niemeyer) Another well-known source, which has influenced the view on RQs as assertions, can be found in the work of Jerry Saddock.1971, who researched RQs calling them queclaratives. Both of these works will, among others, be taken into careful consideration in this thesis. The fact that RQs seem to share more features with declaratives, than with questions can be observed throughout the literature up to this day. A more recent work is for example, Han. 2002, where RQs are transformed to declaratives at post-LF level. RQs are also mentioned together with a) biased questions and b) negative polarity items where the assertive aspect of RQs is described as “rhetorical flavor”(e.g. Guerzoni.2003)

Given the preliminary definition and the further assumptions above, there seems to be an intuitive difference between rhetorical questions and information-seeking questions. Before more sensitive territory is entered, a look at this relative of RQs is the obvious first step. In (2), a) is a RQ while b) is an information-seeking question.(hence ISQ).

(2) a) Who wouldn't like to drive Ferrari? RQ
a') Everybody likes to drive a Ferrari. → underlying assertion
b) Where is the concert hall? ISQ

Although, the difference should be clearly recognizable in this example, the grate between rhetorical questions and information seeking question is rather small. It is assumed, that there are two main factors that influence, whether we perceive a question as RQ or as ISQ: The context and the intonation. Both of them can easily alter the perception of an utterance in the shape of a question. Consider the question in (3) - Already, by changing the natural rising intonation of a question to a falling intonation the questions gains a rhetorical flavor. (Han.2001) This flavor can be strengthened by adding a context, in which cats are out of discussion for the discourse participant. For example at a meeting of allergic people, when someone rather jealously utters (3)a. However, as mentioned above there are various additional strategies to enrich the rhetorical flavor of an utterance, as is exemplified in (3)b. Even without the context introduced above, there should be a clear inclination to understand the utterance as a rhetorical question, with the underlying assertion that nobody likes cats. However, the transition from ISQ to RQ is gradual in this case. The relation of NPIs and RQs will be paid close attention to later.

(3) a) Who likes cats?
b) Who even likes cats?

Let us now turn to rhetorical questions in German. The reason for this is the richness of the German language to mark utterances as rhetorical questions. While it was made clear that, cross-linguistically, the two factors presented above play the main role in detecting rhetorical questions, it was also shown that there seem to be various additional strategies to evoke a rhetorical understanding of a question. For the course of this thesis, linguistic cues that evoke rhetoricity are called *rhetorical markers*.

While some rhetorical markers, seem to be universal, German also offers a number of markers, that are not present in every language: particles. Particles as for example “schon”, “auch” und “denn” have been described thoroughly by Meibauer. In the next paragraph one of the strongest rhetoricity markers in the German language, the particle “schon” will be discussed to create a first subset of RQs that can serve as reference later. We start here, because it is assumed that this certain type of RQ is one of the most distinctive in regards to overt rhetorical markers. At least among German and English RQs..

As it was said above, the particle “schon” is one of the most clear indicators for RQs. It makes it possible to create an overtly distinctive structure which makes it easy to determine, whether an utterance is meant rhetorically. However, one aspect that undermines the explicitness of the particle is that, without context, “schon” is clearly ambiguous between a rhetorical and a temporal reading. (see Meibauer. 1986. p. 114) However, the German language offers tools to dissolve this ambiguity. The following example (4) taken from Meibauer. 1986 illustrates this.

(4) a) Wer holt schon$_{R/T}$ Kohlen?
a') *Who gets PARTICLE$_{R/T}$ coal? / Who is already getting coal?*
b) Wer schon$_{R}$ holt Kohlen?
b')*Who PARTICLE$_{R}$ gets coal?*

As can be seen above, in the case that “schon” precedes the verb, it is incontrovertibly used as a rhetorical marker. The position after the verb makes the sentence ambiguous between a temporal particle and a rhetorical one. While it is possible to explicitly make a sentence rhetorical, there is no way to rule out the possibility of a rhetorical reading completely. In this regard, there is one more important point that needs to be taken into consideration: Rhetorical “schon” can only occur in Wh-questions - questions constructed with an interrogative pronoun (who, where, what, why) - but not in Yes/No - questions. Meibauer here differentiates between endorsement questions and decision questions. (Ergänzungs vs. Entscheidungsfragesätze) This characteristic is exemplified in (5).

(5) a) [Wer mag schon kochen?]$_{R/T}$

a') *Who likes PARTICLE$_R$ cooking? /Who already likes to cook?*

b) [Magst du schon kochen?]$_{T/*R}$

b')*Do you already want to cook? // *You don't like to cook*

One more peculiarity of "schon" shall be mentioned here as a side-note. It seems to evoke a certain effect in assertions, dependent on the content of an assertion. In certain contexts, "schon" seems to reverse the polarity of an assertion. As already mentioned above, this reversal of the polarity is often attributed to RQs as well. The phenomenon is exemplified in (6).

(6) a) S: Ich hab (schon) wieder mein Handy verloren

I have lost my mobile phone again

A: Du bist schon so ein Held.

You are PARTICLE$_R$ such a hero.

S: Ja, ich weiß. Manchmal bin ich echt verpeilt.

Yes, I know. Sometimes I'm really out of it.

However, the effect of "schon" in this case heavily depends on the attribute that the speaker wants to impute to the addressee. Without going into to much detail, it could be argued here, that the particle "schon" rather underlines a certain attitude of the speaker towards the addressee, than to the content of the utterance itself.

Aside from "schon" there is the closely related "auch" which can also evoke rhetoricity in an utterance. It is often replaceable with "schon". For example (5)a above could also be replaced with "Wer mag auch kochen?" This generally happens in questions constructed with the Wh-words "warum(why)" and "how(wie)". Furthermore, "auch" is also ambiguous. However, here the second meaning is not of a temporal nature. The second meaning of "auch" implies that there is at least one other person, who already falls in the set of people that is described by the utterance containing "auch". See (7). The ambiguity of "auch" has the same restrictions as the ambiguity of "schon". So in the rhetorical sense they are interchangeable.

(7) a) S: Magst du auch Kuchen?

a') *Do you also like cake?*

b) Does the addressee belong to the non-empty set of people who like cake?

Furthermore, there is a number of other modal particles which, according to Meibauer - though unable to evoke rhetoricity - have a strengthening effect on the rhetoric flavor of an utterance. The effect of these various modal particles will not be discussed in detail here, but it should be kept in mind, since it will play a part in the second apron of this thesis.

It has been shown above, that "schon" is a strong rhetorical marker. Though, additional markers clearly need to be taken into consideration to accumulate enough information to describe RQs properly. Further strong indicators for RQs will be described before long, so that solid amount of data is available as a base for the forthcoming theoretical and practical work.

The previous paragraph briefly illustrated the unique rhetorical markers in the German language. In the following paragraphs, further surface cues for rhetorical questions will be discussed. Although, the examples still are mostly based on German, the upcoming rhetorical markers can be observed cross-linguistically.

The next indicator for rhetoricity that will be discussed here is negation, which also can play together with the rhetorical marker "schon". Presupposing the example (5) above, Meibauer presents **(8)** to exemplify this relation and the strengthening effect of negation.

(8) a) Wer holt nicht schon$_T$ Kohlen?

a') Who does not PARTICLE$_T$/already get coal?

b) Wer holt schon$_{R/T}$ nicht Kohlen?

b') Who does PARTICLE$_{R/T}$ not get coal?

c) Wer schon$_R$ holt nicht Kohlen?

c') Who PARTICLE$_R$ does not get coal?

Meibauer assumes two different kinds of negations, which can be easily illustrated by replacing the negation particle “nicht” with “keine”. This is only possible b) and c) where a rhetorical reading is possible, not however in a) where only the temporal reading is acceptable. There is however, also a case, where negation is argued to be always rhetorical, namely, in combination with conditional. This is exemplified in (9) which also has been the very first example in this work.

(9) a) Who wouldn’t like to drive Ferrari? RQ

For now, this should suffice to show that negation is a legit tool to evoke rhetorical readings in questions. The topic will be addressed a bit more in the next part, when pragmatic implications of negation in questions are discussed. Similarly, negative polarity items that were also mentioned before can be described better in regard to their semantic and pragmatic impact. Aside from such overt rhetorical markers, there are also other possibilities to detect rhetorical questions. These will be discussed below.

1.2. Tests for RQs

Various properties of RQs can serve as tool to create tests to determine them. In this paragraph (some) of these tests will be laid out, since they are the second major component in preparing a well-rounded research data for the following parts. A detailed elaboration in this field is provided by Saddock, who researched a number of question types in the light of the Speech Act theory. Furthermore, Saddock is often cited throughout the literature, e.g. Han, Caponigra & Sprouse, etc. To begin with, in his work Saddock entitles RQs with the expression queclarative, which might lead to confusion. This means in this section tests for what he calls queclaratives are discussed.

The first example shows two almost similar questions, which both contain a negative polarity item, namely, “a damn thing”. Saddock argues, that the NPI has quite different impacts on the different questions. The first question is apparently of positive polarity, while the second question is of negative polarity.

According to Saddock this seems to make the difference between rhetorical question (or declarativity) and a proper question (interrogativity). At least in combination with the NPI. To prove this point, he shows that there is "differential behavior with respect to two syntactic tests which involve interrogativity or declarativity"

(10) a)Do phonemes have a damn thing to do with language?
b) Don't phonemes have a damn thing to do with language?

First he claims that that only (10)a) can come with the introductory item "after all", which is typical for declaratives, while this is not possible for interrogatives. For this reason, Saddock claims, that 10b) is not compatible with "after all".

(11) a') After all, do phonemes have a damn thing to do with language?
b') *After all, don't phonemes have a damn thing to do with language?

This seems to be a fairly well accepted feature of RQs. However, the author is not completely happy with this test. Though, no further discussion will be lead here, since there was no time to thoroughly discuss these concerns with native speakers of the English language. The next test, that Saddock presents is the linkability of queclaratives(RQs) to assertions via the lexeme "yet". This also seems to be impossible for proper questions, as shown in (12). Furthermore RQs can also be linked to assertions via conjuncts, see (13), while proper questions do not have that ability.

(12) a) Do phonemes have a damn thing to do with language,(?)
yet people continue to believe in them.
b) *Don't phonemes have a damn thing to do with Language, (?)
yet people [don't / continue to] believe in them.

(13) a) After all, isn't Chicago a beautiful city (?)
and besides it's got thirteen Mandarin restaurants

Another syntactic feature that is presented by Saddock does not apply to rhetorical questions but the information-seeking questions. Saddock claims, that only those can connect with the phrase “by any chance”. This is exemplified in (14). However, he does not discuss the semantics behind “by any chance” in detail. Thus, this last “test” has to be handled with care. All in all, it is more suitable to determine, whether a question is an information-seeking question, than, whether it is a rhetorical question.

(14) a) Are the gazanias blooming, by any chance?
b) *Aren’t the gazanias blooming, by any chance?

Though it is not really a test in the sense of the data proposed by Saddock, there is one more important piece of data that has to be kept in mind. It is compactly described by Gressilon. 1990. The main point here is, that RQs have a set of answers different from that of a *plain* information-seeking question. This is illustrated in (15).

(15) a) War das in dieser Situation nicht notwendig?
Wasn’t this necessary in this situation?
- Ganz richtig!
Quite right!
- Ja, natürlich ...
Yes, of course ...

(15’) a) This was necessary, in this situation.
- Yes, (it was necessary)
- (Actually), no, (it wasn’t necessary)

In (15) it becomes clear, that RQs can be answered with, what seems like an acknowledgement to the underlying assertion of the RQ. Keep in mind however, that the natural answers shown in (15’) are acceptable as well. Furthermore, they also are compatible with a plain assertion as well. Before this data is concluded, first assume (16).

(16) Gehst du mit? Are you coming (with us)?
a) - *Richtig. *Right
b) - Ja. Yes

Here it is assumed, that the answer a) seems to be a bit weird at least. However, Gressilon didn't pay this data enough attention. It is quite easy, to create a context, where a) wouldn't seem so wrong after all. To achieve this goal, the question in (16) only needs to be biased. Assuming that it is more likely that the addressee is coming along, than not, (16)b) doesn't seem so inappropriate anymore. This data shows, on the one hand, another clue, why one may assume, that RQs are covertly assertion. On the other hand, it provides the insight, that it is most likely necessary to pay close attention to the possibilities of answering questions which makes it more similar to *true* (biased) questions.

In conclusion, this section has shown, that although rhetorical questions can be identified by quite a number of surface (structural) cues and syntactical tests that follow the right intuitions, it is necessary to pay more attention to the semantic and pragmatic aspects of RQs. The set of RQs we can identify, grew with the advancement of this sections. Though, it was shown that observing these surface cues from a descriptive point of view is not enough. To cover the set of all rhetorical questions, it is inevitable to take a look at the semantics and pragmatics of rhetorical questions.

2. Semantics and pragmatics

In the previous section we have seen that there is quiet a number of factors that need to be paid attention to, to describe RQs accurately. Before this challenges can be addressed, first it is necessary to decide whether RQs should be treated as hidden assertions, or as a special kind of questions. To begin with, the hypothesis, that RQs are a type of question rather than a type of assertion will be established. To successfully defend it, a theory coherent with the data from the previous chapter as well as the research in the field of semantics and pragmatics presented in this chapter, will be proposed. The first step to elaborate this theory is to describe, what it means for an utterance to be a question in contrast to an assertion. After that, rhetorical questions will be observed from various points of views. The main part here will be the discussion of the idea by Caponigro & Sprouse which will serve as a base for the pragmatic analysis of RQs presented in this work. To sum this section up, RQs will be differentiated from other closely related question types in terms of their semantic and pragmatic analysis.

2.1. Semantics of questions

Though there are various theories to describe the semantics of questions, the underlying ideas are more or less similar throughout the literature. For the sake of simplicity the relatively old work of Hamblin will be used as an example here. Keep in mind that this work is also an important cornerstone for most more recent theories about the semantics of questions. The main idea is that questions, in contrast to declaratives are represented by a set of propositions rather than one single proposition. This works for Yes/No-Questions as well as Wh-Questions. (17) illustrates both cases:

(17) D_e = {Anna, Beth, Caroline, …}

a) Does Anna want coffee?

Semantic representation: {(that) Anna wants coffee, (that) Anna does not want coffee}

b) Who wants coffee?

Semantic Representation: {(that) Anna wants coffee, (that) Beth wants coffee, (that) Caroline wants coffee …}

= {p: ∃x [p = that x (is a person and) wants coffee]}

The case of Yes/No questions should be intuitively clear: There are only two answers to the question whether Anna wants coffee. Either she wants coffee, which also can be mapped to the answer "Yes" or she doesn't want coffee which maps to the answer "no". The transformation operations should be more or less trivial and not discussed here.(Hamblin.1973) In case of the Wh-question, the set of propositions depends on the underlying set of entities. For example questions created with "who" are mapped to all entities in D_e that are also persons. Respectively "where" is mapped to places and "what" is mapped to things. The question word can than be represented with the help of an existential quantifier, as can be seen in (17)b.

When is obviously not mapped to the set of entities, but should work in a similar way and will not be further explained here.[1] The second special case of Wh-questions are questions build with why. For now it is proposed to assume that they work with a set of proposition that cause the proposition in question. In (18) p is an answer to q, q contains the proposition in question and r describes the reason for q. If the condition in p holds, it is a possible true answer, else it is not.[2]

(18) a) {p: ∃p [∃q [∃r [p = that if r then q]]]}

b) - Why is the street wet?

- Because it rained.

q = that the street ist wet.

r = that it rained

p = it rains -> street is wet

1 It is assumed, that although the basic idea of Hamblin might work here, the elaboration would not be worth its time here, since further explanation in regard to tense would be necessary. Let's just assume, "when" works with a set of points in time.

2 The author is not capable of presenting a source for this derivation. Please let me know, if this is worked out somewhere, or if the approach is questionable to begin with.

Please keep in mind, that the derivation proposed for Why-questions is only tentative.. All in all, this should be enough to understand the basic functionality of the semantics of questions. Further approaches to derive the meaning of questions build on this simple way.

2.2. Rhetorical questions as assertions

Let us now turn to the first commonly discussed semantic interpretation of RQs by Chung-hye Han. 2001. Although Han uses a somewhat more sophisticated approach to the analysis of questions, the main idea, as already mentioned above, is quite similar. The approach to the semantics of questions used by Han is based on a work by Groenendijk and Stokhof and is described as a theory "suggest[ing] an algebraic account of the possible values of wh-words"(p.11. Han.2001) The core difference lies in the set that describes the possible answers. While the representation of Yes/No-questions is in the end the same as seen above in Hamblins approach - a set containing a negative polarity and a positive polarity mapped to positive and negative answers respectively - Wh-questions are represented by a power set. A power set "is a six tuple $(B,1,0,\cap,\cup,')$, where B is the domain of the algebra, 0 and 1 are elements of B, corresponding to the empty set and the unit set respectively, $\cap$ and $\cup$ are binary functions corresponding to intersection and union, and ' is a unary function corresponding to the complement."(p.211. Han.2011) The power set ultimately describes a set containing each possible set of answers. Each set can be mapped to one possible world in intensional semantics. Only one of the given sets can be mapped to the actual world, namely, the one with the true answer in that world.
Han describes, that the derivation of rhetorical questions is based on the properties of this set. While for plain information-seeking questions the possible values for the Wh-element are denoted by the power set described above, in rhetorical questions "wh-phrase ranges over only one possible value, namely the bottom element, which corresponds to the negative quantifier. And so the question returns a partition with a single block [-or a set containing only one element-] the denotation of which is equivalent to an assertion."(p.221. Han. 2001) With this description Han explains two common phenomena of RQs. The assertive character as well as the (often) reversed polarity. However, how is this explanation motivated? This question is addressed in moderate detail below.
The basic idea is that, although the semantics of a question stays as described above, the pragmatics of

Wh-questions force a modification of the LF(Logical Form) of the question. As a possible explanation Han lays out the Gricean maxims. To be more precise, Han refers to "the first part of the Gricean maxim of Quantity"(p. 215. Han):

(19) Make your contribution as informative as is required (Grice. 1975)

As Han describes, she derives her notion of informativeness "from various probabilistic ways of modeling epistemic states"(p.215. Han.2001) The definition given by Han is: "The notion of 'informativeness' [is] relative to the individual's degree of belief in a certain proposition p in a given context c."(p.215. Han.2001) This in effect means, that propositions, that a certain listener believes not to be true in the context are more informative, than propositions that the listener already beliefs to be true. However, only if the listener believes the speaker to be truthful. In this case, the listener would be forced to update/change his beliefs and thus the proposition can be seen as informative.(see Han. 2001) For questions this means, that "when a speaker is formulating a question to find out whether p or not p, for example, s/he formulates the question in the form of the proposition that would be most informative if it turned out to be true.[…]In other words, the likelihood that a speaker will use a negative question not p? is equal to the speaker's assessment of the probability of p" (p.215. Han.2001) This is illustrated in (21).

(20) a) S: Does it rain? → S believes it does not rain.
b) S: Doesn't it rain? → S believes it does rain.

Semantically both propositions, *(yes), it does rain*, and *(no) it does not rain* are given when posing a Yes/No question. However, as described above only one of the answers "is consistent with the pragmatics of Yes/No questions[, namely the answer with opposite polarity.] Thus, the [answer of opposite polarity] is selected as the assertion expressed by [a] rhetorical Yes/No question. In effect, rhetorical Yes/no questions implicate the speaker's expectation towards the answer in the strongest possible form"(p.216. Han.2001) This means, "the speaker's expectation towards the negative answer is asserted as the speaker's belief."(p.216. Han.2001).

A similar explanation can be given for Wh-questions as well, according to Han, since they also seem to

have “implications in terms of speaker’s expectations towards the answer.”(p.216. Han.2001) To be more precise, in certain conditions, the speakers expectation is, that the amount of propositions that serve as true answers is smaller, than the amount of proposition, that are no true answers to the question. This can be, for example, caused by focusing the main verb in positive Wh-questions. Remarkably, this effect seems to be given in Wh-questions with negative polarity. (see Han.2001) This is illustrated in Hans’ example (46) and (47) here depicted as(21).

(21) a) Who didn’t finish the paper?
b) **Speaker’s expectation:** Most people finished the paper.
a') Who FINISHed the paper?
b') **Speaker’s expectation:** Most people did not finish the paper?

The bias effect, that is encountered here, seems to be exceptionally strong in RQs, or rather enforced by the Speaker. This means, for Han, that the speaker forces the question into the shape of an assertion by his or her expectations. A similar effect seems to be evoked by NPIs. In the work of Guerzoni and van Rooy it is shown, that the speaker’s expectation towards an answer, or the understanding of a question as a RQ can also be described in different ways. At least the subset of RQs that uses NPIs as rhetorical marker. It is assumed that the approaches presented above can be roughly mapped to Han's definition of informativity and the result to pose questions with a certain bias. However, the approaches differ in so far that in questions with NPIs the rhetorical flavor can be traced back to certain lexical items, the negative polarity items. We already saw examples for NPIs before.

(22) a) Who even likes cats?
b) Do phonemes have a damn thing to do with language?

There are arguments that there is a certain strength to NPIs. However, Guerzoni.2004 mentions in a footnote (no. 12) that the strength effect itself can hardly be seen as explanation for bias in questions because it is unclear how this should work. (see Guerzoni.2004). This also means, that the even stronger effect, the effect to rhetoricity, may not be connected to strength either. However, this point is left open for discussion.

2.3. Rhetorical questions as proper questions

Guerzoni was already mentioned, but there is another similar account provided by van Rooy. The main point that can be found throughout both of these works is the denotation of minimal as well as maximal endpoints on contextual scales by certain lexical elements. In the case of NPIs this would be the low endpoint of a scale. (Rooy.2003) The main point that can be deduced from Guerzoni and van Rooy is, "that the rhetorical force should result from the limited way in which the question is still information-seeking"(p.255. Rooy. 2003) Below it will be illustrated briefly, how Guerzoni and van Rooy account for this limitations with the help of scales and the effects of NPIs on these scales.

As already mentioned above, Guerzoni as well as van Rooy are discussing the functionality of NPIs. Their claims about rhetorical questions are more of a side-product of their work. One of the main points that both of the researchers are discussing are the presuppositions that are evoked by NPIs. These presuppositions are what effects the answer-ability of a question - whether they evoke a biased or a rhetorical reading. Guerzoni argues that these presuppositions are carried by, a possibly hidden, even. Hidden in the sense, that the same presuppositions hold for strong NPIs, even if no "even" is available. (called even-type NPIs in Rooy.2003) . The presuppositions of even are the following according to Guerzoni. 2003:

(23) a) hardP = p is the least likely proposition among the alternatives.
b) easyP = p is th most likely proposition among the alternatives.

One of the presuppositions, namely, the easyP is responsible for bias and rhetorical readings in questions. Before we pay attention to this certain presupposition, first the scales that NPIs point at need to be explained a bit more. Said scales have a high and a low endpoint and refer to contextual information about the theme of an utterance/question. Guerzoni generally uses the example of test questions, which can be scaled from easiest to hardest problem. Applying the presuppositions introduced above, the hardP would be mapped to the most difficult problem. The easyP is mapped to the easiest problem respectively. It is not clear between Guerzoni and Rooy as what triggers those presuppositions. (Rooy. 2003) However, it is not necessary to explore this problem at this point.

The question is, how a point on a scale is related to bias and rhetoricity. For this purpose Guerzoni's example is explored.(Guerzoni.2004. (18))

(24) Can Sue even solve [Problem 2] ? 'ambiguous'

a) <Problem 2, Problem 5, Problem 3; . . . ; the easiest problem>

ScalarP: For any alternative x, it is MORE likely that S can solve x than that Sue can solve Problem 2. p is the LEAST likely among the alternatives. **(hardP)**

b) <the most difficult problem, Problem 3, Problem 5; . . . ; Problem 2>

ScalarP: For any salient alternative x to Problem 2 it is LESS likely that Sue can solve x than that Sue can solve Problem 2. i.e. p is the MOST likely among alternatives. **(easyP)**

The argument that b) evokes a rhetorical or biased reading can be traced back to the probability of the answer. If the answer is very likely to be "yes", than the most informative and expected answer would be "no". At this point, we are at a similar state as in Han's work that was presented before. All in all, it can be seen, that all the accounts can be traced back to one thing. The speaker's expectation towards the answer to the question. This will be taken into consideration later.

It has been shown that the semantics and pragmatics of rhetorical questions can be evaluated, based on pragmatic principles that can be either derived from the structure of the utterance, the intonation, or presuppositions of lexemes. However, the approaches presented above seem to only partially solve the riddle posed by RQs. Recent accounts seem to adopt an approach that is based on the common ground theory to describe the pragmatics of RQs. The main idea is, that various types of questions can be described by different states of the proposition in question in the common ground. Truckenbrodt.2004, first argued, that the goal of RQs is, to evoke a certain answer in the addressee, by making it reconstructible for him. This seems reasonable, but a bit imprecise. For that reason, first another proposal with a similar foundation will be presented here: The approach devised by Caponigro & Sprouse. To describe RQs in the common ground properly, they used a modified account of common ground analysis along the lines of Gunglogson, which is known for her work on questions in the shape of assertions. In the same manner as in Gunglogsons work, the CG presented here consists of three sets of propositions. The speakers beliefs (SB) the addressees beliefs (SA) and the CG, which contains,

what is mutually believed by the speaker and the addressee. The last set can only contain propositions from the intersection of both of the speakers beliefs. However, not all proposition in the intersection have to be in the common ground. (Gunglogson. 2001) Based on this the following definition for ordinary questions and RQs is proposed:

(25) a) **Definition of OQs (ordinary questions) by Caponigra & Sprouse 2007**
An OQ is an interrogative clause whose answer is not known to the Speaker, but the Speaker thinks the Addressee may know it. An answer is required in order for the dialogue to be felicitous. Only the Addressee can answer.

b) **Definition of RQs by Caponigra & Sprouse. 2007**
A RQ is an interrogative clause whose answer is known to the Speaker and the Addressee, and they both also know that the other knows the answer as well. An answer is not required, but possible. Either the Speaker or the Addressee can answer.

Furthermore, Caponigra and Sprouse use their CG-analysis to differentiate RQs from other question types in the same manner as Truckenbrodt. For this they use the various constellations of the information state in the discourse participants beliefs. This will be discussed at the end of this discussion, since it is assumed that Caponigras and Sprouse's account on RQs needs to be modified first.

2.4. Rhetorical questions in the common ground

In the previous section it was shown, that the Common Ground analysis proposes a promising tool the describe RQs as proper questions. However, in this section, a modification of Caponigra & Sprouse's work will be proposed and it will be backed up with some new examples. Furthermore, the approach that will be shown here, is also - at least in an abstract way - compatible with previous works, such as Han, van Rooy and Guerzoni. Although, the latter were already addressed briefly by Caponigro & Sprouse. It will also be shown, that the account proposed here, goes in line with the attribute of

persuasiveness, which is generally ascribed to RQs, but not so much discussed in the semantic and pragmatic literature. (Caponigra & Sprouse. 2007) To enter this section, let us again look on their definition for RQs repeated in (26).

(26) a) **Definition of RQs by Caponigra & Sprouse. 2007**
A RQ is an interrogative clause whose answer is known to the Speaker and the Addressee, and they both also know that the other knows the answer as well. An answer is not required, but possible. Either the Speaker or the Addressee can answer.

The present thesis accepts the premise, that "an answer is not required, but possible" and that both can answer. However, a revision of the first part of this explanation for RQs is deemed necessary. The first aspect that needs to be questioned, is the stance of the speaker towards the (complete true) answer of the RQ. It is by no means necessary, that the speaker needs to believe in the truthfulness of the proposition he wants to be conveyed by the RQ. Take for example (27).

(27) S: Wer mag schon Spinat?
Who likes PARTICLE$_R$ spinach?
A: Richtig, lass uns zu McDonalds gehen.
Right, let's go to McDonalds.

In this scenario, the speaker wants to persuade the addressee to go eat fast food instead of something healthy. In fact in this example, none of the discourse participants needs to believe in the proposition conveyed by the Speakers RQ. However, this will not make a difference, as will be shown shortly.
For the purpose of persuasion, the speaker does not have to be committed to the proposition that nobody likes spinach. He may even like spinach himself. In fact, the only one who needs commit to this answer to the RQ is the addressee. This means, if we adapt the CG analysis from Gunglogson and Caponigra, the only element - if anything - that needs to be in the Speakers belief SB is the proposition, that the addressee can be convinced to believe that nobody likes spinach. By not contradicting the RQ or by explicitly answering it with an appropriate answer, the addressee then makes two changes to the CG and the participants beliefs. First, the proposition, that he believes, that nobody likes spinach is

added to the common ground. Thus, it is also in SB. Second, the believe, that both, speaker and addressee believe in the truthfulness of the proposition, that nobody likes spinach is added to the addressees belief. This allows a proposition to be added to the common ground with out the speaker committing to it. Furthermore, he gets additional information about the state of mind of the addressee, meaning the addressees believes. The whole process is illustrated in (28).

(28) a.) 1. the speaker poses a question Q.
2. The addressee finds an answer p (which is the answer intended by S)
3. The addressee assumes, it is accepted by the speaker (based on the structure of the question)
4. The common ground is updated: A believes p.
5. The addressees believe is updated: A believes that S believes p.
6. A believes that S & A believe p & S believes, that A believes that S & A believe p.

The procedure shows that although p is in the common ground, it is perceived differently by the Speaker and the addressee. The fact, that the addressee updates the CG can also be taken as further enforcement of the believe in p of the addressee. Thus, A may commit stronger to p, than if it was an utterance instead of an RQ. Furthermore, A thinks that he is not alone with his opinion as to the truthfulness of p. From this thought process the persuasiveness of RQs can be derived. The approach proposed above also explains the answer-ability of RQs. If a RQ would be interpreted as an assertion, one would assume a rhetorical question to violate the Gricean cooperative principles. This is argued by defenders of the Speech Act theory, who claim, that a RQ underlies some kind of indirect speech act, thus a conventional implicature, which in return is evoked by violation of the Gricean principles. (Meibauer.1986, Saddock.1974) In the light of the present proposal, it is more likely that the genuine question character of RQs does support the Gricean maxims. The speaker intends to convince the addressee of a proposition, which he may not believe (Maxim of quality). However, the speaker has no obligation to accept it at all. In fact, it is the addressees task to accept or refuse the proposition that is discussed. Despite, the fact that the speaker has certain rather strong expectations, he still leaves an option for the addressee, which could be described as cooperative behavior. This also can be linked to

the Maxim of Manner. Furthermore, the maxim of quantity would be violated, if a RQ would be interpreted as assertion, since the shape of the assertion would suggest that a set of proposition is given, instead of a single one. By pointing out the most informative answer with help of the structure of the RQ, the speaker utters as proposition that is desired in the CG, without disregarding the possibility that this might not reflect the addressees believes. The Speaker gives the addressee just enough information to figure his point of view out. To sum this up, the attitude of the speaker, is more convincing or persuasive, since he is not forcing his point of view on the addressee and manipulates him into believing that speaker and addressee are on an equal stance in regards to a certain proposition. Furthermore, since the speaker has uttered a question, the addressee commits to the proposition, since he understands it as the answer, to the question, that he has found.
Although, there can not be given any evidence, it seems natural, that even in long debates, people often base they're counter arguments on RQs posed during the talk of the opposition. Maybe, an analysis of data in the light of this elaboration would be helpful.

It was already shown, by Caponigro & Sprouse, that RQs actually behave like "ordinary questions" in respect to their answer-ability. However, one could argue, that there are minor differences. The next paragraph will elaborate on that. To begin with, it is assumed, that the possibilities to answer RQs are more comparable with possibilities offered by biased questions, which differ from plain information-seeking questions, as was shown by van Rooy and Guerzoni. The shift in the probability of an answer to be a true answer, also affects the response possibilities, or rather methods. For example it seems quite natural, that biased questions can be answered with agreement, instead of a "true" answer. The answers in (29) refer to both possible readings of the question along the lines of Guerzoni.2003.

(29) a) S:Did Kim even solve Problem 2?
b) A: Right, she is not the brightest, huh?
b')A: ?Right, she is really smart!

Furthermore, it seems natural, to mark answers in a certain way, when they are not in accord with the expectation of the speaker. In English, such answers are often introduced with "well" or "actually". In German there is a similar option with "eigentlich". This assumption is exemplified in (30).

(30) a) S: Wer mag schon Spinat?
Who likes PARTICLE$_R$ spinach?
A: (Naja), eigentlich mag ich Spinat schon ziemlich.
INTERJECTION, actually/well, I quite like spinach.

This peculiarity can be also mapped to RQs as seen above. With this answer the addressee accepts, that his answer might not be what the speaker expected. Although, it would also be possible for the addressee to strongly commit against the stance of the speaker by directly answering the RQ. However, in this case it seems natural to focus the subject, which still hints some kind of idiosyncrasy of the answer. It is assumed here, that this focus has an effect similar to "well" described above, but with a different attitude towards the speaker.

(31) a) S: Wer mag schon Spinat?
Who likes PARTICLE$_R$ spinach?
A: ICH mag Spinat.
***I** like spinach.*

The next paragraph briefly discusses, the connections of the proposal given here to previous accounts on RQs. To begin with, how can this approach be connected with the work of Han.2001? Although, it cannot be taken as evidence for the transformation at post LF-Level of questions described by Han, it can be easily be linked to the reasoning behind her work. To be more precise, to her interpretation of the Gricean maxim of quantity and her definition of informativeness and its implications. As Han elaborated, questions are normally constructed with a certain expectation towards the answer based on the principle of informativity. If we presuppose, that the speaker is aware of these expectations he is capable of exploiting them, which is assumed to happen in RQs. By constructing the question in a way, that makes the proposition, that the speaker desires the addressee to believe the most informative, the speaker evokes an effect of achievement in the addressee by finding the answer that was intended by the speaker to begin with. Similarly, a speaker could utilize the presuppositions evoked by NPIs to reach his goal, a question with a rhetorical reading. Questions like "After, all who gave birth to you"

evoke a rhetorical reading, simply by posing a question with only one true answer. All in all, the approach presented above does not consider various sub sets of RQ differentiated by the means they are evoked with, but rather tries to capture the RQs in their entirety. In the next paragraph, it will be worked out, how the approach presented above allows to distinguish RQs from other related question types.

2.5. Relatives of rhetorical questions

To begin with the relatives of RQs will be distinguished by observing the information state in the Common ground and the participants believes. The following table is derived from assertions made by Caponigra & Sprouse. However, it will soon be shown, that this table is oversimplified.

(32)

	Speaker knows p	**Speaker does know p**
Addressee knows p	RQ(?), Examination question	OQ
Addressee does not know p	Examination question	Questions with no answers

The first row should follow naturally from the definitions given by Caponigra & Sprouse. The second row can easily be traced back to the work of Truckenbrodt.2004. The case of examination questions is special, since the speaker does not know whether the addressee knows the answer or not(he only expects it), so both combinations are possible for examination questions. The importance in both cases is, that there is no knowledge about the answer in the common ground which differentiates examination questions from RQs. Questions with no answers, however, should be self-explanatory. Questions, that simply can not be answered because neither speaker nor addressee are capable of knowing the answer. Keep in mind, that there are ordinary questions, where neither one knows the answer based on the circumstances. This happens, when neither person knows the answer, but there is no mutual belief about it. It should be clear by now, that the typology given by Caponigro and Sprouse is insufficient, or at least incomplete. To be more precise. There should be at least six possible combinations with the

properties shown above alone. It is not taken in account in this table (32) how the question-type changes, depending on the state of the common ground. Is the answer in it, or not? Truckenbrodt.2004, although also relying on a common ground analysis, describes the question-types introduced above differently.

Below Truckenbrodts assumptions in regard to the common ground are presented. However, Truckenbrodt describes what changes in the common ground, rather than its state while the question is posed. So “ordinary” ISQs aim to update the Speakers believe, while RQs and other types of questions want to update the addressees belief, if we describe Truckenbrodts assumptions in terms of Gunglogson's account.

(33)[3] Demand to update common ground can primarily aim at:

a. S knows whether p	speech act question (OQ)
b. A knows whether p	pedagogical questions
	rhetorical questions
	monologue questions
c. S weiß, dass A weiß, ob p	Examination questions

Pedagogical questions and monologue questions are described as a similar phenomenon in this account. However, observed from the point of view of the Caponigra & Sprouse approach, more fine-grained differences can be derived.

Monologue questions are, as the name suggests, questions that are posed without expecting an answer from the addressee. This does not mean, that the addressee knows the answer. The answer is given by the speaker generally shortly after. As a result, the question is more of a reversed information-seeking question. The speaker assumes a question, that the addressee might have, poses it himself and instantly answers it. This, for example, may be the case in lectures. Headlines are also put in this category by Truckenbrodt, since the article answers the question. (Truckenbrodt.2004) This means the speaker presupposes/believes that the addressee does not know the answer to the question. However, since this type of question is mostly addressed to large audiences, this does not necessarily have to be the case.

3 Example from Truckenbrodt.2004 – translated by the author.

Pedagogical questions are questions that are posed in way that makes the answer clear, which makes them quite similar to rhetorical questions. However, they are used in different contexts. Pedagogical question have a rather educational flavor and, although this has not been formally researched it is assumed, that they are invoked with different markers, than rhetorical questions. One marker that comes to mind for pedagogical questions would be the word “really”. For example: “Was it really necessary to buy so many shoes?” The context should be self-explanatory. It is argued that this kind of question presupposes that the addressee is already somewhat aware of the right answer to the question. (Koshi.2002) In this respect this type of question might fit better into the definition of RQs proposed by Caponigro & Sprouse. If we translate Truckenbrodts assertions into the table presented above, the result would be (34). However, there are still open questions, which shows, that there might be other levels that need to be taken into consideration to differentiate the question types presented here. It has been shown in the discussion of Caponigro & Sprouse's approach, that the intention of the speaker play an important role. Furthermore, the common ground can be decomposed in more detail, by taking in account the speaker's expectations towards the answer in the form of his beliefs about its information state in the addressees beliefs. Especially, when the common ground itself is actually empty in regards to the question, the speakers and addressees beliefs need to be analyzed in more detail.

(34)

	Speaker knows p	**Speaker does know p**
Addressee knows p	RQ(?) Examination question, monologue question, pedagogical questions	OQ, monologue question(followed by answer)
Addressee does not know p	Examination question, pedagogical questions(?)	Deliberative question

2.6. Conclusion

In conclusion, the approach by Caponigra & Sprouse was brought closer to the proposal made in Truckenbrodts work from 2004. Modifying the former account with the latter account as a guideline seems to provide a feasible account to analyze rhetorical questions. Since it seems to describe the information state of the discourse participants, as well as the necessity of the shape of a question better, than the original account by Caponigra & Sprouse. Truckenbrodts account was modified in so far, that the speakers information state was taken into consideration in more detail. In general, it was shown, that it seems necessary, to pay more attention to the actual raison d'être of rhetorical questions. It was argued, that RQs are not simply assertions coated in the shape of a question, but rather a proper type of question that is used as a diplomatic tool to confirm or dismiss certain information states, rather than forcing new information on the addressee. Furthermore, it was shown, that, from this, a certain consensus between speaker and addressee can be derived, that may have a positive psychological effect, which strengthens the attribute of persuasiveness.

3. Research - Study at the long night of sciences

The following part will be the link between the theoretical and the practical part of this work. By using RQs manually extracted from a Corpus consisting of German Tweets - postings on a social-media platform - the ability to detect RQs with little to no context information was researched. For this purpose the impact of certain linguistic cues on the interpretation of an utterance was passively and actively observed with help of a survey. The study was conducted at the long night of sciences at the University of Konstanz. In addition to the strategies to detect RQs the survey also provides data, on how linguistic intricacies are perceived and described by the general public.

3.1. The survey - Method and procedure

To test how the detection of RQs a survey was developed. The first part of the survey consisted of 50 items which were rated in regard to their rhetoricity. In the second part of the survey participants were asked to give a definition of RQs or a description on how they determined whether an item was rather rhetorical or rather not. The next paragraph will elaborate on first task of the survey.
The study was conducted at the long night of sciences at the University of Constance. An event, where most of the faculties may present their work to bring science closer to the general public. The survey was part of a presentation about rhetorical questions as a preliminary study.

(35)

	Frage	**Bewertung** **1 = überhaupt nicht rhetorisch** (verlangt eine Antwort/Information) **10 = eindeutig rhetorisch**									
		1	2	3	4	5	6	7	8	9	10
1	Wer räumt denn schon sein Zimmer auf?	O	O	O	O	O	O	O	O	O	O

Method: The survey was designed as an offline-study were the participants had to fill out a questionnaire, that was split in three parts. The first part contained a brief explanation about the task, as well as some back ground information. Two main factors were given here. The research of the - at first glance - contextually poor nature of tweets on Twitter and the note, that successfully exploring rhetorical questions might help to develop more sophisticated systems for sentiment analysis in the foreseeable future.

The explanation of the task contained three core points. First, the participants were asked to keep the background, namely social media and Twitter, in mind. Second, they were asked to rate the questions as varied as possible. This was done to prime participants to carefully think about how strong they would rate certain phenomena. Third, the second task was mentioned, where the test persons were asked to give an explanation or a definition on how they differentiated the various items.

Participants: As there were no restrictions, as to who wants to take part at the activities at the long night of sciences, the participants of the survey were wildly mixed. Almost no demographic information was collected. However, information about the first language was collected, since it was considered the most important factor for the ability to detect RQs or linguistic intricacies in general. However, a major part of the 41 participants at the long night of sciences had declared German as their first language. Only a handful declared otherwise(two people). Moreover it was ensured, that people had no knowledge about rhetorical questions or research about RQs to find out, how such linguistic phenomena are perceived in general. Although, this might decrease the value of the study for RQs per se, it is assumed that this gives improved results in regards to the natural detection methods one uses to differentiate linguistic properties. As a control group a relatively small number of six people with more or less extensive knowledge about RQs were also asked to take the survey. All of them were native Germans and part of them also has conducted scientific research about RQs.

Test items: The items were designed after an in-depth analysis of a corpus of 187 tweets ending in question marks.[4] The main criterion for a RQ was the assertive force in the sentence. To research linguistic cues, as seen in the sections before, various rhetoricity markers were constructed into the items. The two main factors were particles and polarity items, furthermore negation played an

4 The available corpus grew over the course of the work. In the end approximately 1000 tweets were analysed for this thesis.

important role. Additionally question types were mixed. In total the survey contained 22 Wh-questions and 28 Yes/no questions. Alternative questions were not tested. Furthermore punctuation was varied to research its linguistic effects. In general, excessively punctuated utterances were assumed to be perceived as more rhetorical. Finally, some tag-questions were included since they have been argued to be close relatives to RQs. (Saddock.1971)

To determine the strength of the effects tested here, control items were constructed. These were plain information seeking questions but partly modified with special attributes, similar to the ones introduced above, to determine, how fine grained people differentiate the items. All items were thoroughly discussed and judged by the author and a co-worker.[5]

Procedure: As mentioned above, the study was part of a presentation at the Long night of sciences. Before taking part in the survey, people got a short introduction to the topic. They were not told a definition for RQs, however, some examples of RQs were given. After the introduction the visitors of the presentation were asked to fill out the survey. For this task they only had a short time span of approximately 10 minutes. After that, the people were asked to participate another study, where they were given a explicit definition for RQs. Two people took part in the second study first. Their results were included in the data for the control group. After taking part in both studies, the participants were given a small or a big reward (candy) depending on the performance in the second study.

3.2. The survey - Results and discussion

To begin with some flaws in the survey need to be considered. Despite the fact, that it was designed to explore natural processing strategies to detect the fine differences in the presented items, the conceptual formulation might have misleading effects. To be more precise, the problem lies in the fact that people were asked to differentiate between various levels of rhetoricity and had no other variable at hand to differentiate the items in the first task of the questionnaire. This may lead to situations, were people rate items more (or less) rhetoric, for the lack of a better attribute. The likelihood of such situations is not small, since, as was shown before, defining the scope of RQs is not a simple task.

5 We also received support from other researchers at the University of Constance, which (informally) helped modifying the survey.

Furthermore it was difficult to test the influence of certain constructions separately. The innate property of language to use various strategies to disambiguate certain utterances - often at the same time - can be named as a reason here. Resulting from this - the complex interaction between certain factors - it is difficult to draw explicit conclusions from the results presented here. Furthermore and even more critical is the fact, that even the impacts of well-defined aspects are difficult to process for this reason.
These critical points make it impossible or rather futile to use the findings of this survey as reference for the second part of this work on a statistical level. However, the author believes that tentative clues on how to detect rhetorical questions can be deduced from the work conducted here. The next paragraph will elaborate on this. It needs to be kept in mind, that the second task of this survey will support the investigation of the rating-task of this study.

That being said, the first observable feature is, that there seems to be no significant differences between the data of the participants and the data of the control group (t-test > 0.65). This suggests, that the general idea about what a rhetorical question is seems to be quite accurate. However, one observation that is mentioned here, only backed up by estimating, is that the control group tended to be more extreme in regards to the rating. This in return means, that the participants of the original study distributed exceptionally high or low ratings rather hesitantly.
Nonetheless, except for a handful exceptions the results coincide with the expectations under which this survey was developed. The 21 items that were expected to be rated highly rhetoric have a mean of 7.28 with seven elements exceeding the standard deviation.
It is eye-catching that all items containing a rhetorical “schon” have extremely high ratings above 8 with ten being definitely rhetoric. This seems to prove the fact, that the rhetorical “schon” seems to be an explicit marker for rhetoricity in German. It needs to be said here, that it was attempted to avoid possible ambiguities with the temporal “schon”. However, even when both readings were available theoretically, the rhetorical reading has been expected to be more prevalent in the respective items. This is shown in (36)

(36) a) Wer räumt denn schon sein Zimmer auf? 8.659
b) Wer kann das schon wissen? 8.317
c) Wer braucht das schon zu lesen? 8.098

Tag-question that were also assumed to obtain fairly high ratings, have been rated very differently depending on the phrasing of the tag-part of the question. Since there is quite some variance of possibilities in German, this might be associated with the semantic or pragmatic differences in these phrases which could not be researched or explored in detail in this work. In (37) the differing rating of the various tag-questions is exemplified.

(37) a) Wer hier ständig um Ansehen bettelt, hat doch kein stolz mehr, oder? 7.56
b) Du begleitest mich doch, ne? 5.76
c) Du weißt doch wie wir hier wieder herauskommen, nicht wahr?! 4.3

Example a) is puzzling since it was not assumed to be rated rhetorical at all. Yet it has the highest rating among the five tag-questions in the item list. It seems, now is the time to turn to the analysis of the second part of the survey. How can this puzzling observation be explained in the light of the analysis of the second task? Here, it becomes clear, that some participants assume, that some of the close relatives of RQs (partly) discussed in the theoretical part are actually types of rhetorical questions. For example some statements can be lead back on exclamative or deliberative questions.[6] For example (38).

(38) a) "manche sind keine Fragen, sondern Ausrufe".
a') *"Some are not questions, but exclamations"* => **exclamative**
b) "Überlegung ob man darauf antworten könnte."
b') *"Considerations whether an answer is possible"* => **deliberative**
c) "Eine rhetorische Frage verlangt nicht eindeutig eine Antwort. […] (z.B. Warum regnet es heute?)"
c') *"A rhetorical question does not definitely require a response. […] (e.G. Why does it rain today?)"* => **deliberative**

It is conspicuous, though, that a large number of participants seems to rely on the assertive character of rhetorical questions to detect them. It is even more eye-catching, that some test persons even described

6 Deliberative questions are questions that convey the feeling of saying out loud what one thinks. Questions introduced b "I wonder..." belong to this kind. Truckenbrodt.2004. Exclamative questions are exclamatives in the shape of a question. For example, the commentator of a football game can utter: "What a goal!" d'Avis. 2001

the peculiarity of some RQs to reverse the polarity of the assertion they try to convey. This is exemplified in (39).

(39) a) “Dinge über die allgemeiner Konsenz herrscht, werden ins Gegenteil verkehrt[…]”
a') *“Things that are commonly accepted are reversed to the opposite”*
b) “Gegenteil ist als Aussage gemeint.”
b') “*Opposite is meant as assertion.”*
c) “Fragen[,] die ohne direkte Antwort eine Information liefern”
c') *“Questions, that without a direct question, deliver an answer.”*
d) “Sprecher hat ein gewisses Vorurteil im Bezug auf die erfragte Antwort”
d') *“Speaker has a certain prejudice in regard to the answer of the question”*

The examples above do not only describe the assertive aspect of RQs, but show that other commonly researched factors are also considered when detecting RQs. For example statement a) can be seen as a paraphrase for the analysis of the common ground. C) also can be investigated in this respect. As was pointed out, the addressee is asked to commit to the information conveyed by a RQ or, in other words, make it prevalent in the common ground. The last example d) shows the relation of RQs to biased questions. As we have seen, strategies that are used to describe biased questions, also can be mapped to RQs regarding their grammaticality. (Guerzoni.2003)

Overall, although the results did not provide any all to surprising explorations, it has been shown, that the generally accepted features of RQs are valid and it should be possible to translate at least some of them into some kind of machine processing system. The question whether this process results in a feasible possibility for detecting RQs is still in question. A semantic and pragmatic analysis is likely to be necessary for the task of detecting RQs. This concludes the present section. In the following part the data collected here is put to use and some technical details, especially about Twitter, that were omitted in this section especially, are described.

II. Practical Part

1. Challenges of automatic processing of data from Twitter

The automatic processing of linguistic intricacies is an intelligible desire in the field of computational linguistics. Furthermore the steady development of the electronic - especially social - media urges researchers to look for more sophisticated approaches to language analysis. Most notably, the art of processing questions in social media is a growing field lately. (e.g. Dent & Paul. 2011, Li et al. 2011, Paul. 2011, Zhao & Mei. 2013 etc.) The second part of this work will present some preliminary work in automatic detection of rhetorical questions. Raw-data will be extracted from the German Twitter-stream. The technical details of this extraction are explained in the next section. Consequently two challenges will be discussed in the course of the practical part:

(1) a) Detection of rhetorical Questions in text form.
b) Processing of linguistic data from social media/Twitter

The difficulty of the first challenge can already be deduced from the first part of this work. As it has been shown, even a theoretical analysis of RQs is dependent on a number of factors which subtly change the strategies to determine whether an utterance is a RQ.
To smooth out the most obvious problem when processing data from Twitter, the lack of phonological information will be discussed briefly. We have seen before, that there are more or less distinct points of views, when it comes to the intonation of RQs. This is exemplified in the work of Han, which, as was seen before, describes, that RQs, in contrary to ISQs, have a falling intonation more similar to that of an assertion, than that of a *true* question. Accepting that theory, it should still be clear, that this should not be a decisive factor, to abolish the idea of this practical part. This is based on the findings from the previous section. It was shown, that people were by all means capable of detecting RQs without this *crucial* feature. Furthermore, we have seen in the theoretical part, that there is quite some variance in the strategies to construct and mark RQs. This means in return, that there should be various ways to determine RQs.
This is the bright spot, that will be used for this preliminary work: The richness of various structural linguistic cues in the German language. As was seen before, the German language in particular offers various strategies using linguistic attributes visible in the surface form of a sentence. This means, there

are more or less certain rhetorical markers in the phrasing of RQs. In the last part of the previous section, it was discussed, how these cues are perceived and what strategies people use to identify RQs with little to no information.

Although not exclusively, it was shown that structural cues were perceived as indicators of varying strength for RQs. Following this, the present work tries to remodel some of the strategies people used in the survey into a script to detect, or help detecting RQs. The script is additionally backed up by the theoretic findings from the first part of this work. In (2) the main elements, that are used as rhetorical markers are summarized.

(2) Table of rhetorical markers in German

Rhetorical marker	*Example*
Rhetorical particles	**Wer mag *schon* Spinat?**
Strengthening particles	Wer hat denn eine Million Euro?
NPIs (weak vs. **strong**)	Wie kann ich ihm *jemals* wieder in die Augen schauen?
Negation + conditional	**Wer *hätte nicht* gerne einen Ferrari?**
Negation	Sind wir *nicht* alle ein bisschen Bluna?
PPs	Was *zum Teufel* kann so viel Lärm machen?

Observing this table, it becomes clear, that the rhetorical markers that actually evoke rhetoricity (bold in the table) are rather small in comparison with those, that strengthen rhetoricity. This finding is also backed up by Jörg Meibauer, who also differentiated between rhetoricity evoking and rhetoricity enforcing cues as was seen in the previous section. For this reason, a script that simply sorts Tweets in the two categories *rhetorical* and *non-rhetorical* is assumed to provide unsatisfactory results. To make the best out of data that has been collected so far, it was decided to create a script, that rather than detecting RQs, sorts the input according to the likelihood of the items to be rhetorical based on structural cues. The likelihood of rhetoricity was judged in a rather informal manner for the lack of adequate data. A rating system, that is partly based on the findings from the study presented before and partly based on theoretical claims and intuition was implemented. Expecting a likelihood between 0 and 1, (3) illustrates the expected procedure to be executed by the script.

(3) a) Wer hat denn jemals Spinat gemocht? Rating: 0.5
b) denn-modifier: x 1.5 Rating: 0.75
c) jemals-modifier: x 1.25 **Rating: 0.9**
→ it is very likely that a) is a rhetorical question.

The example above is only tentative to illustrate the intended functionality of the script presented in the next section. The likelihood for a sentence to be rhetorical, henceforth its *rhetoricity rating***,** is determined by the modifiers it gets. The system checks, whether certain rhetorical markers match with the sentence that is currently processed. The technical details as well as the evaluation of such a system is discussed in the next section. Before that, another challenge needs to be addressed first.

Talking about the second challenge, the most important thing is, that social media greatly influences the development/use of language. This compulsion for language change can easily be described with the helping example of Twitter, which is also used as an example for the work at hand.
Twitter is a social media platform designed to share small pieces of information with the general public. For this purpose, posts are restricted to 140 characters each. The influence of this restriction lead to an adaption of language and the delivery of a new text-form called Micro-Text by K. Dent & S. Paul, who have adapted it from Rosa & Ellen 2009. Derived from this, Twitter can be called a Micro-blogging service. (Dent. 2011) The properties are summarized

(4) *The properties of Micro-Text are the first part in the challenge of processing data from social media: "micro-text is short, often consisting of a single sentence or even a single word. Second, the grammar used is informal and unstructured, and there are often abbreviations and errors.(Rosa and Ellen. 2009)"* K.Dent & S. Paul. 2011

The author of this thesis assumes, that not all these properties are caused by Twitter, but are owed to a number of communication options offered throughout the Internet, e.g. chat-rooms, especially in connection with online games and forums. The point here is, that the Internet in general heavily affects language change/development. In this regard some findings of the present work, should not be regarded solely in the light of Twitter, but rather in the light of language change under the influence of the

Internet.

Irrespective of the cause for the language change, this shift has to be taken into account, when trying to process data from Twitter. This has already be shown in a number of works on questions in Twitter, as for example, Dent & Paul.2011 quoted above. Despite the fact, that their attention was in the favor of information-seeking questions, the challenges posed by the language on Twitter stay the same for this work. For this reason the next section will elaborate in detail, how the challenge of language change due to the Internet is addressed, before the actual linguistic analysis takes place.

2. The Script

In this section it will be illustrated how the challenges mentioned above were addressed. In the first paragraph utilized resources will be described. This is followed by a few words about the raw-data that is intended to be processed by the script. Subsequently the processing of the data will be explained, which will be divided into two sections. The first paragraph will lay out how the data was prepared for analysis considering the linguistic difficulties posed by micro-text. In the line of this task these challenges will of course be discussed in detail. In the following paragraphs the search mechanisms for RQs and the rating system are described. To sum up this section, a evaluation of the script presented here will be given.

2.1. Resources

As explained before, for this thesis a script was written that should be able to detect, or at least help detecting RQs in Twitter. It seems natural to explore third-party tools before developing a new system. In the course of this project, a number of third-party tools were combined with unique search mechanisms and the rating system mentioned before. The technical design parameters as well as the third-party tools that were used, are described below.

The whole work was executed on an Ubuntu 12.04 TLS 64bit version based on Unix. The script was written in Python, which was pre-installed on this system.. Choosing a (interpreted) scripting language for this task seemed natural, since they are mainly used for this kind of corpus analysis. Another option

would have been to use a object-oriented programming language, e.g. Java. However, this idea was castaway with the thought, that object-oriented programming or at least a simulation of it, is also available for the "most popular" scripting languages (Python, PERL, Ruby …). Furthermore, despite the fact, that Java is platform universal, it wasn't used since the program is expected to be more depended on the raw data it needs to process, than the platform it is running on. Furthermore, Unix based systems **seem** to be the norm anyway. Other kinds of programming languages were not even considered. Among the number of (interpreted) scripting languages, Python was chosen since it was deemed the most comprehensible and clear programming language. Python is an (interpreted) scripting language. This means, the code does not have to be compiled into a set of instructions for the machine, but is rather processed by an interpreter, who can execute instructions in real time. This makes the development more flexible, since applications can be programmed at runtime (Ousterhout. 1998) In the official Python documentation it is described as followed[7]:

(5) *Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.*

As mentioned above, Python offers the possibility to use object-oriented programming, which - in the eyes of the author - has a few benefits. Most importantly it helps to keep the code clean and more well-ordered. Aside from that, the reasons mentioned above, as well as personal preferences lead to the decision to use Python. For this thesis Python V. 2.7x was used since some of the third-party resources used during the further procedure were not compatible with V. 3.x yet.

Regarding third-party NLP-resources used in this project, the most important and probably also most well-known, is the library 'Natural Language Tool Kit - NLTK' for python. The NLTK provides, as its name suggests, access to a large number of interfaces for various NLP-resources, such as WordNet or the Stanford NLP group.[8] The following paragraphs describe the basic functionalities of the third-party tools used for this script. Later, the way they are implemented into the script will be illustrated.

7 https://docs.python.org/2/tutorial/index.html(last accessed: 18.08.2014)
8 http://www.nltk.org/(last accessed: 17.06.2014)

PoS-Tagger: The PoS-Tagger is provided by the Stanford NLP group and offers tagging-functions, as well as the possibility to train it on a certain corpus. For this work however an existing corpus was used: The Negra-Corpus provided by the Univeristät des Saarlands. The corpus consists of roughly 60.000 tokens and was originally created at the University of Stuttgart, where the tokens were annotated with PoS-Tags. It was additionally annotated with syntactic structures. The base for this corpus lies in German newspaper texts taken from the Frankfurter Rundschau.[9] This means the underlying language is most likely much less colloquial than the language used on Twitter. Despite this small inconsistency, the Negra-corpus was deemed adequate enough for the preliminary status of this work. The composition of a fully appropriate corpus would break the scope of this work, since it would not only require to manually tag a large amount of data, but also because of the special nature of language on Twitter, which was already addressed briefly and will be illustrated in more detail below.

Sentence-tokenizer: The tokenizer can be found in the nltk.tokenize.punkt module which belongs to the NLTK project. "[It] divides a text into a list of sentences, by using an unsupervised algorithm to build a model for abbreviations words, collocations, and words that start sentences."[10] The sentence tokenizer is fairly sophisticated and performs reasonably well after cleaning the data as discussed below. Instead of training it, an existing training file, german.pickle, from the punkt model, downloadable via the nltk.download(), function in python was used. As can be read in the Readme-file, the training file for the tokenizer is based on the *Neue Zürcher Zeitung AG Neue Zürcher Zeitung,* a Swiss newspaper.[11]

Stop words: The NLTK library provides a corpus of stop words for various languages, which is used for language detection which is described briefly in the following paragraph. Unfortunately, it could not be made out, from what data the German stop word corpus provided by the NLTK is composed.

9 http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html (last accessed: 18.08.2014)
10 http://www.nltk.org/_modules/nltk/tokenize/punkt.html(last accessed: 09.08.2014)
11 Accessable after download via nltk downloader in ~/nltk.data/tokenizers/punkt/README

2.2. Input – Raw-data processed by the script

The raw-data fed to the script consisted of Tweets extracted from the open German Twitter-Stream via the Twitter-API. It needs to be noted here, that it is not possible to access collections of old tweets, so all tweets were downloaded from the stream in real time during a certain amount of time. In the case of this work, data was provided by the faculty of computer sciences of the University of Constance which collaborates with the linguistic faculty in the frame of a research project, related to this thesis.
The raw-data was only preprocessed by the faculty of computer sciences in the sense, that only tweets containing at least one question mark were extracted from the Twitter-Stream. The data consisted of text files containing varying numbers of Tweets, with one Tweet per line. The files were encoded in UTF-8 and were also processed in that format. A last noticeable remark is, that, although the data was downloaded from the German Twitter-Stream, a not unremarkable amount of Tweets was not composed in German. There are various factors explaining this, like bilinguals, visitors from other countries etc. It is also connected to the way, the search engine for certain tweets, provided by the Twitter API, works.[12]

However, it is no challenge to sort out these Tweets. For this purpose a function was implemented that identifies the language of a Tweet via a corpus of German stop words provided by the NLTK. To summarize its functionality it compares whether the amount of German stop words is larger than the amount of non-German stop words in relation to the number of words in the tweet. If so the tweet is German, else it is in some other language, which does not play a role for the script. Non-German tweets were simply thrown out. This first step of manipulating the data leads us to the next section of this part.

12 https://dev.twitter.com/docs/api/1/get/search(last accessed 15.08.2014)

2.3. Preprocessing of the data

As mentioned a few times before, the language used on Twitter needs to be differentiated from other written text forms. As a result of the nature of micro-text and general Internet-jargon, it is desirable for the raw data to be *cleaned* to enhance processability by the NLP-tools seen before. To clean the data collected from Twitter, help was offered from K. Dent and S. Paul who wrote a sophisticated tool for micro-text analysis. However, they specialized on the English language, which made an direct adaption rather difficult. Yet some functions were translated into the script at hand to clean the raw data. With their work in mind, the process of the cleaning process can be divided into three sub-tasks.

(6) I) processing Twitter-specific objects: Addressee, Retweet-marks
II) interpret or purge Internet-peculiarities: emoticons, links, hash-tags
III) interpret or purge linguistic flaws: spelling, abbreviations, (punctuation)

The first point is not a particular difficult problem, since these objects have a fixed position in a tweet and are uniquely marked. This allows for an easy detection and removal with the help of regular expressions. To be more precise, in the script, each tweet is transformed into a list of tokens, by simply using the Pythons split() function, which splits up strings of characters at spaces. Then it was checked, whether the first object in the list was a RT marker, or an addressee and if so, it was removed. In the latter case, since more than one addressee is possible, the script repeats the process until there is an object, that does not resemble an addressee. In (7) the bold elements represent, what the script removes.

(7) a) addressee: **@userA @userB** check this out …
b) Retweet-marker: **RT** Some interesting news …
c) **RT @userC** What a funny story …

The present script also saves the information whether the Tweet is a retweet or addressed to a certain person, which might help determining the context. To describe this more precisely, one might deduce from this information, whether a tweet is posted in a context more similar to discourse or if the writer is rather talking to the world in general. If anything, this is a rather bold theory, though.

However the storage of this data also helps sorting and categorizing the tweets if the user desires to do so. Although the Twitter-Specific objects do not pose to much of a challenge, the following points of the list are a different story. Some of them do not only hinder NLP-resources at working efficiently, they also may convey semantic or rather pragmatic information which is fairly difficult to interpret on an automatic level. This will be elaborated in more detail below, but first the obvious and explicit problems will be discussed:

Hash-tags: Hash-tags can appear in two forms. Either at the end of a tweet to describe/summarize the content of the tweet, or they can be used as a replacement for a word or phrase inside a sentence. Hash-tags are used to categorize tweets by linking them together to a collection of tweets that can be quickly accessed. (Dent & Paul. 2011) The possibilities of forming a hash-tags are more or less infinite. The only existing conventions are a hash at the beginning of the hash-tag (giving it its name) and the necessity to omit spaces. They are normally written in camel-case when describing a phrase, which is not a convention however.

Initially used as meta-data, they also are used to replace topic-positions in phrases. Thereby, the hash-tag does not always fit exactly in a grammatical sense. A way to address this problem would be, to try whether a sentence is grammatical, if the hash in the hash-tag is omitted. However, some hash-tags may also replace phrases that are necessary for the grammaticality of the sentence. To derive these phrases into a grammatical sentence from a hash-tag is a more complex task, that would need much more resources than could be provided for this work

Though, hash-tags can be removed completely. This is fairly easy to implement with the help of regular expressions. Most hash-tags should be detectable by this rather simple regular expression for them:

(8) a) Hash-tag = “ #[^\s]+ ”

This expression already should produce a very minimal amount of mismatches, though, this cannot be proven statistically at this point. Further handling of the problem is currently not of great importance, due to the shallowness of this approach, with which the present script works. Since deep grammar analysis is not used at all, the number of problems, hash-tags evoke is reduced and is just discussed here for consistency.

Web-links: Similar problems occur with web links. Although mostly separated via punctuation, they sometimes can occur in the middle of a sentence. In this case it is even more difficult to derive a processable item. For this reason they are simply deleted in the script at hand, since they have no apparent value for language research. Together with the assumption, that the sentence tokenizer, mentioned above is capable of differentiating dots in web-links from punctuation marks, this should minimize potential problems. There might be more elegant ways to address the problems posed by hash-tags and links, but for this work, what has been presented here, should be enough. However, this is still just a small part of the challenges posed by the language on Twitter. One problem, the previous paragraph was already hinting, is the use of punctuation on Twitter.

Punctuation: It can be observed on Twitter and throughout various other communication methods relying solely on written language, that punctuation is frequently used very excessively. The linguistic importance of this seems to be barely researched so far. The survey presented in the previous part of this thesis provided some first intuitions. Abide from the implications for the linguistic meta-data, this quirk also affects the functionality of the script. In more detail, the performance of the sentence tokenizer is impaired. It is not designed to deal with multiple punctuation marks following one another, which leads to a large amount of data waste, since it detects a (empty) sentence for every punctuation mark that directly follows another. As a result, this derogates the performance of the whole script. Although not critically damaging the functionality, this problem should be addressed. This was done by implementing a function that, in a first step, replaced each occurrence of multiple identical punctuation marks with a single one of the same type. In a second step accumulations of several different punctuation marks were replaced by the very first one in the string. See (9) for examples.[13] [14]

(9) a) ??? —> ?, !!! —> !, … —> .
b) ??! —> ?, !!?! —> !, ..?. —> .

13 The last type of punctuation mark accumulation might not be solved ideally and the precision of the detection of question-shaped sentences might suffer, but the use of punctuation is fairly inaccurate on Twitter to begin with, so some loss or another has to be taken for granted at this stage.

14 The new version 3.0 of the NLTK can deal with multiple punctuation marks, which would make this part obsolete. However, it would require moving the whole script to Python and NLTK 3.0 which was rejected at this time.

With a similar method other punctuation marks, that are unnecessary for the analysis are taken out. This includes all kinds of brackets and parenthesis as well as various forms of quotation marks that might pose a problem for further processing.

The problem with some punctuation marks is the "meta-linguistic" or pragmatic effect that was first pointed out in the survey in the previous part of this thesis. Though, barely verified, these not-so-well-understood phenomena are part of what drives the present work to begin with. This means simply ignoring/deleting this certain characteristic is not acceptable. For this reason, despite of using the cleaned tweets for further linguistic inquiry, the original tweets are kept, in an attempt to put the aforementioned information to good use in separate functions for this "meta-linguistic" data.

Another phenomenon that is relying heavily on punctuation is the use of emoticons. As the name suggests, those are mostly used to represent ones emotional stance. In the authors opinion their importance as pragmatic indicators is more questionable as the importance of accumulated punctuation discussed before. For this reason emoticons are simply omitted in the clean version of the tweet and ignored in the original version. As often done before, they are sorted out by using a regular expression. However, the invention of emoticons seems to be in a productive state of development, which makes it difficult to clean them out completely. Paired with the emoticons there is one other method to express emotions, or even perform virtual actions, namely the bracketing with asterisks. This is described in Dent.2011 as an attempt to dramatize the content. It is exemplified in (10).

(10) a) My guinea pig just died. *cries*

b) Today's my birthday! *puts on party hat and dances around*

As illustrated above, the use of asterisk-bracketing is also productive and can be quite complex. For the lack of means to analyze this kind of "meta-linguistic" information, it is also cleaned out with the help of a regular expression in the clean version of the tweets. From here one last step needs to be taken. The step into the field of acronyms, which can be divided into two groups. The first group consisting of acronyms used to simply cut certain phrases short to safe space and the second group consisting of abbreviations that are used to express an emotional outburst - normally varying degrees of amusement.

(11) a) See you = cu, gl = good luck, jk = just kidding …

b) laughing out loud = lol, rolling on floor laughing = rofl, what the fuck = wtf …

Since the usage of abbreviations is also more or less productive, it is fairly difficult to deal with them. Although the most popular ones could be cleaned out, it was decided that in this shallow approach, abbreviations are mainly ignored, since they, just as the two points before, barely affect the performance of the overall-functionality and seem to have little or no exploitable pragmatic value whatsoever at this stage.

A last point of the preprocessing is the general deviation from spelling and grammar in Internet-slang that was mentioned as a general property of micro-text. This problem is really difficult to address in the scope of this work. As was shown in Dent.2011, a thorough analysis using finite-state morphology and a large lexicon is necessary to successfully account for these problems. The system developed by K. Dent and S. Paul is capable of replacing the most common standard spelling errors or deviations with the correct spelling. This also includes some spelling peculiarities of Internet-slang, that might convey linguistic value of some sort. For example the elongation of certain words by means of repeated letters, or the change or drop of letters in words. The latter probably having less linguistic meaning, than the former. However, it is not possible to elaborate in the course of this work. Additionally, using homophones is a common mean to abbreviate words. (see K. Dent. 2011) As mentioned above, in the end, all these varieties of abbreviations were simply ignored.

Unintentional spelling mistakes are also not taken into consideration. Contrarily, it seems like unintentional spelling mistakes are more or less uncommon. This can be explained by the existence of spell checkers and auto-correct features, especially on mobile phones, which are often used as social-media devices. However, grammatical errors are not corrected by this. This can be ignored due to the fact, that the script is not intended to execute a deep grammar analysis in the first place.

Although, most of the examples presented in this paragraph are in English, it needs to be kept in mind, that quite a number of them was adapted in German as well. More so, there seem to be certain trends in all of the peculiarity discussed above. Some are only used by certain demographic groups, others change over time. To sum it up, they are also subject to language change and are not yet formalized. This means that some strategies - or to be more precise regular expressions - used in this script may become obsolete, or at least inaccurate, in the future.

As a final task in the state of preprocessing, all data is stored in an object that saves various pieces of information to analyze each tweet properly.

(12) **Tweet:** The original tweet
Cleantweet: The tweet after preprocessing
Questions: Each sentence in the tweet that ends with a question mark
Phrases: Each sentence in the tweet that does not end with a question mark.
Booleans: Information about Retweet-marks and addressees saved as booleans (true/false)

The first two pieces of the object were elaborated on above. The condition for a sentence to be treated as question is for it to end in a question mark. Similarly phrases are all sentences that do not end in a question mark - or simply put, all other sentences. At this state of the script, these sentences do not play a roll but the option to work with them should be held open. The data about retweet-marks and addressees is saved as boolean, a truth value. Primarily, to provide the possibility to sort the data.

2.4. Detection mechanisms and the rating system

As already pointed out, for language analysis the present script implements instances of the Stanford PoS-Tagger as well as the NLTK Punkt Sentence Tokenizer and accesses a NLTK Corpus consisting of stop words for language detection. These resources are all made available by the NLTK library and are used in combination with some searching mechanisms developed for this script. To begin with, a bi-gram and trigram analysis was conducted. With the theoretic claims in mind, that RQs have the same overt structure as information-seeking questions, this was only seen as a test run, whether particles in German make any difference at all. However, this analysis did not yield any fruits and thus wasn't pursued any further. Furthermore, previous work has shown, that a rule-based approach was rather successful, at least for information-seeking questions (K.Dent, S.Paul. 2011). As a pre-study for rule-based RQ detection approaches, the present script uses regular expression, as well as PoS-tags to judge sentences in tweets. It has already been shown, that there are factors that are definite signs for RQs (e.g.

“rhetoric schon”, strong NPIs, negation with conditional) as well as factors that only strengthen the perceived rhetoricity of a sentence (other particles, weak NPIs, negation, punctuation etc). Based on the variety of these cues and the differing influences on the rhetoricity of a sentence, it was attempted to turn this script into a rating system, that rates the potential of tweets and sentences in tweets, to contain or be a rhetorical question. The idea is also based on the survey, that was conducted in the study presented above.

The first step to implementing such a rating system, was to work out, how the structures listed above, and more, could be detected in tweets, without relying on deep grammar analysis. The first obvious idea was the use of regular expressions. However, this approach poses a number of problems, which can be summarized in the problems with the “rhetoric schon” alone. As a reminder, “schon” is ambiguous between a rhetoric reading and a temporal reading. It is quite difficult to model a regular expression, that can differentiate between “temporal schon” and “rhetorical schon”. Nonetheless, a extensible dictionary was implemented to provide a well arranged possibility to search for regular expressions. For the dictionary a function was implemented that can be applied to strings and searches it for a regular expression with help of a key to the dictionary. This can for example be used, to search for strong NPIs or expressive phrases like “what the fuck”.

To address the problem posed by “schon”, the use of a PoS-tagger was assumed necessary. Since the plain n-gram analysis with (and without, see above) PoS-tags was unsuccessful, an other way to work with structural cues was implemented. The first step was, to implement a function, that searches for loose tag-chains. A loose tag-chain is a list of tags, that can be found in the tagged sentence in the order of the tag-chain without regard to the distance between the single tags. Tag-chains were also written into a dictionary and processed similarly to regular expression. The functionality of loose tag-chains is exemplified below.

(13) a) Wer zum Teufel mag schon rhetorische Fragen?

b) [(Wer,interrogative_pronoun), (zum,prep), (Teufel,noun), (mag,verb),…]

c) [”interrogative_pronoun”,”verb”,”noun”]

d) match: Wer, mag, Fragen

In (13) a) represents the original phrase that is compared with the tag-chain. However, b) is, what is given to the function: a list of word-tag pairs saved in tuples. The second parameter, that is given to the function is c), the tag-chain. The function iterates through the list of word-tag pairs(tuples) until it finds a tuple, where the first element of the tag-chain matches the tag in the tuple. The function then enters the recursion loop, where the first element of the tag-chain is cut off. Furthermore, all elements including the matching element are cut off from the list of tuples. This means both lists shrink with each recursive call of the function. The function ultimately returns true, if the last element of the tag-chain list matches a tag in the word-tag pair list. It returns false, if the last element of the tuple-list is reached without a match. By cutting off both lists, mistaken matches of tags and tagged elements are prevented. For example “noun” in c) cannot match with the noun “Teufel”, since it is already cut off at the point when the tag “noun” is processed. So it can only match with the noun “Fragen”, as is intended. During the development, this function was expanded to also process regular expression for single words as well. This means, tag chains as in (14) could also be processed.

(14) a) [”interrogative_pronoun”,”verb”,”schon”,”noun”]

Elements are now matched with both, the PoS-tag and the word itself in each word-tag pair. This is possible due to the uniqueness of the tags provided by the negra-corpus. To put it differently, there is a near zero probability for the script to successfully match a PoS-tag with a word, except the input contains something were spelling has gone terribly wrong. For this reason, the example given above also matches successfully with (13)a in the same manner as described above. The only difference is, that two match-conditions are tested each time the script compares a tag from the tag-chain with a tuple from the list of tuples representing tagged sentence. This final version of the function, together with the function to search for regular expressions, does the main work of the script, aside from the cleaning.
The function described above is nothing groundbreaking and the more difficult task is, to find the appropriate tag-chains to find structures that possibly indicate a rhetorical question. The main concern on a technical level that needs to be kept in mind is the occurrence of heavily nested sentences. For simple sentences, the tag-chains should work decently well. Processing embedded RQs, however, needs some very carefully assembled tag-chains. While sentences introduced by subordinating conjunctions are still manageable since they are tagged as such, things are getting more complicated,

when these structures are omitted. Fortunately, the STTS(Stuttgart-Tübingen Tag-Set, based on the negra-corpus) is highly distinctive, which makes it possible, to distinguish sentences by the number of predicates and possible tags for topics(subject/object). Despite these considerations, one might still come up with various flaws in this system. However, for the preliminary state of this work, no further steps were pursued in improving the present script substantially.

The *rating-system* that is based on the searching mechanisms presented above is located in a separate main body inside the script. This makes it possible, to fine-tune the program, without actually changing essential functions of the program. This also means, the script presented here is rather easily adaptable to other tasks in the field of corpus-linguistics.[15] As already mentioned before, the rating system is based heavily on statements made in the literature and is backed up by study conducted for this thesis. Resulting from this the table in (15) presents the rating criteria considered in this script and its impact on the rating.[16]

(15) Table of the rating criteria based on the rhetorical markers in German

Positive rating criteria	*Impact on rating*
Rhetorical particles	Very strong x 2.0
NPIs[17]	Weak x 1.5
Strengthening particles	Weak x.1.25
Negation	Weak x 1.25
Question type (Wh-question)	Very weak x 1.125
PPs	Very weak x 1.125
Punctuation	Very weak x 1.125

Basically it is assumed, that the initial rating for each question is 0.5 with 1 being definitely rhetoric and 0 being not rhetoric. This was decided, because it was assumed there should be no bias, before any testing was done. When the rating of an element reaches 1 or goes beyond, then further testing is

15 During the development of this thesis it was also considered to use the script for detection of verbs with auxiliary particles.

16 How the criteria was translated into loose tag-chains/regular expressions can be observed in the analyzer_dicts.py class. Everything should be sufficiently commentated.

17 The NPIs used here were classified in Lichte.2005

stopped. Values are rounded up to one digit after the decimal point. Opposites of the criteria presented above, in particular positive questions and yes/no-questions get a penalty of 0.75. Again, the values proposed here can be fine-tuned at will and are only used for a tentative run testing the overall functionality of the script.

Prepositional phrases and Punctuation were assigned a very weak modifier since although it was claimed that they would strengthen rhetorical readings, their actual impact could not be attested sufficiently. The rhetorical “schon” und “auch” were declared as the only modifier to definitely evoke rhetoricity in a question. The former only in Wh-questions. It is not tested in yes/no questions. Unfortunately, it is not possible to search for conditional-structures. However, it is arguable, whether negated conditional structures are explicit rhetorical markers in yes/no questions. All in all it already becomes clear here, that much further work has to be done. For this reason, the conclusion here is: There are many more intricacies that could be taken into consideration and tested in the script, however the configuration presented above was taken as the final one to be presented in this thesis.

2.5. Evaluation

It has already been hinted, that the script written for this thesis can be polished in several ways. However, when one carefully assembles the loose tag-chains for the search mechanisms as well as the regular expressions the tool at least supports the detection of RQs.. Unfortunately the assumption that the lack of phonological and contextual information would force Twitter-users to use more explicit strategies to mark rhetorical questions couldn't be confirmed universally. Compared to a manual search, there is a large number of Tweets that fall through the net, when the script is applied and even then the rating can only be taken as clue, not however as evidence. This means it is impossible for the program to provide a list of rhetorical questions extracted from an arbitrarily large number of Tweets. However, the searching and rating-mechanisms work as intended and although the script is not yet capable of finding RQs itself, it helps to find the direction in which one must go to accomplish the task of automatically extracting rhetorical questions. This further path is discussed in the following and last section of this work.

3. Future possibilities - Thought experiments

In this section the practical part of this work will be concluded. The practical as well as the theoretical research of this work has shown, that a shallow approaches to the automatic extraction of RQs bear more difficulties than benefits. Nonetheless the results of this work are used here, to provide a view into the future. Some more or less realizable strategies will be shown. Necessary resources as well as the feasibility of the approaches will be elaborated and discussed.

3.1. Implementation of sophisticated grammar systems – XLE

The first idea discussed here will be the usage of a sophisticated grammar analysis system. To be more precise the use of a Lexical Functional Grammar (short:LFG). The author emanates the use of the XLE-platform for LFG processing. However, since the considerations here are of a purely theoretical nature, other grammar-processing platforms and other grammar systems (HPSG, TAG, etc).
The first and easiest example, that comes to mind, because it could not be addressed before, is the detection of the construction *conditional + negation.* The flexibility of XLE-grammars easily allows to detect constructions of this type. Without going in to much detail it should be possible to create a new statement type, *rhetorical*, which requires both of these properties as well as the shape of a question to be triggered.

(16) a) Wer hätte nicht gern einen Ferrari?
a') Who would not like to have a Ferrari?

Maybe an in-depth analysis of negation alone would help to detect certain kinds of sentences with rhetorical flavour. As we have seen in the first part of this work, negation undeniably is a valid method to create RQs. Since it is possible, to get various interpretations of a single sentence via XLE, it might be possible to provide a plain-question reading as well as a rhetorical reading for sentences where both readings are possible. Considering that, it might even be easier not to mark rhetorical sentences with a statement-type *rhetorical,* but to make rhetoricity a separate property. At this point no explicit proposal as how to implement this, is made, since it would go beyond the scope of this chapter. However, one

other point that would be profit from the use of a grammar-system is addressed.

The detection of particles. As mentioned before the particle “schon” in German is an ambiguous marker for rhetorical questions. However it can be disambiguated so that the use of the particle can be clearly divided into a temporal “schon” and a rhetorical “schon”, where both are licensed by different grammatical structures. (Maibauer.1986) However, with the examples here, there is still a problem similar to the one in the shallow approach. Though, these tools might be suited better to detect such surface cues, there are still a number of RQs that are purely context driven. We have seen such RQs before, for example:

(17) a) After all, who gave birth to you?

This should already show, that a grammar-system alone is not enough to detect every kind of RQ. However such grammar-systems can be linked to a semantic system. Semantic systems will be discussed in the next paragraph.

However, despite the possibilities there is another rather big downside to the use of such system: The robustness. LFGs generally rely on accurate punctuation, lexicon and grammar. (Butt et al.1999) It can already be foreseen, that many peculiarities of Internet-slang will hurt at least one of these aspects, especially the lexicon. Though, the tendency to violate grammar constraints also needs consideration. This would greatly reduce the feasibility on platforms like Twitter, since punctuation, as well as grammar and spelling all suffer in this environment. At least without a proper reassessment of the raw-data.

All in all, the approaches presented here only provide a partial solution, since there is still a big amount of RQs not relying on the approaches presented here, even if one would implement the interpretation of NPIs as well. Furthermore, more attention would have to be paid to clean the data for use with a grammar-system. A system as described in K.Dent & S.Paul.2011 would at least be necessary to process data from Twitter properly in XLE. It has already been stated, that this approach would benefit from the implementation of a semantical system, so thoughts about semantic systems will be addressed next.

3.2. Implementation of semantic systems

By now it should be clear, that some kind of semantic and pragmatic understanding is necessary to determine rhetorical questions. To address this problem the use of a semantic system, especially ontologies is proposed. In theory, these constructs could cover another problem with the detection of RQs. How this could work and what practical requirements would be necessary, is discussed below.

First of all, it should be illustrated, what is understood as an Ontology in this thesis. As described in Jurafsky & Martin.2009 an ontology is a "hierarchical organization[…] that captures the subset/superset relation among […] categories"(p.606. Jurafsky & Martin.2009). As exemplified in their restaurant example(p.611), Ontologies can also represent scales, as in this case the cost of a restaurant. If we map this to rhetorical questions as in (18), a machine could deduce from the ontology, that McDonalds is a cheap restaurant and thus the question could be subject to restraints along the line of Guerzonis approach to biased questions and RQs.

(18) After all, who likes to eat at McDonalds?

Despite the fact, that it would be possible to model some kind of context with this approach, there are still some downsides. To begin with, the explanation here is rather one dimensional. It is not difficult to assume a large number of contexts, were such scales are not part of the interpretation at all. For example, if a group does not like an otherwise decent restaurant for personal reason, or because of bad experiences, then one could still utter a RQ along the lines of (18) to convince the group, that the restaurant is not a good choice to eat at. Adding personal preferences to an Ontology, despite not being completely impossible - probably - would be a ridiculously complex task, which would only make sense in a highly developed artificial intelligence with learning ability.[18]

When talking about automatic language processing systems, one might also assume the implementation

18 Assuming a system like SIRI, Alice or other common Artificial Intelligence applications, one could imagine a memory for such preferences being implemented to a knowledge base, a storage of information about a certain world or situation. (see Jurafsky & Martin.2009) By furthermore enhancing the knowledge base with an ontology, as illustrated above, one might cover a few more instances of RQs.

of a probabilistic system. It was shown in van Rooys work (Rooy.2003), how rhetorical questions (together with biased questions, again) might be represented by the means of the probability distribution of the truthfulness in the set of answers. However, here the question, how to teach an automatic system to judge such probabilities, is very difficult to answer. One might imagine possibilities, using things such as analysis of search engine results, when encountering questions with explicit topics, like the restaurant question above in (18). However, questions as in (19) would still be almost impossible to detect.[19]

(19) a) Was kannst du (überhaupt)?
b) What can you do / Can you do anything at all?

Despite all that, It needs to be kept in mind, that judging such questions, based solely on text form, is only based on "wild guesses" for humans as well. More so, when speaking about data on Twitter, where the context is extremely restricted.

As a last remark, the topic of sentiment analysis is addressed in the following paragraph. The popular definition, that RQs express a certain assertion of the polarity opposite to that of the question (e.g Han.2001) has been discussed before. Although, data speaking against such an approach was provided, these kind of RQs are still a valid subset of RQs in general. One of the initiating thought processes of this thesis, that regrettably could not be addressed further, was the idea to exploit this polarity phenomenon to develop an approach on sentiment analysis, that is based on (deep) linguistic analysis.

However, in the course of this thesis, the question may rise naturally, whether it might be more lucrative, to use this very sentiment analysis, to help detecting RQs. This might be connectible to the probability approach, presented above. By analyzing the sentiment that is prevalent in e.g. the Twitter-Community, it might be possible to detect, whether the phrase in question is a RQ, or not. This would require two steps. Finding out, what the most informative answer, or the underlying assertion of the RQ is, based on its structure, and second, analyzing all tweets that can be related to that assertion or topic by some kind of semantic search in regard to the sentiment they convey. This is exemplified in (20).

19 However, semantic analysis of the particle "überhaupt" might do the trick. Though, the particle can be omitted.

(20) a) Immer dieser furchtbare Verkehr. Ehrlich mal, wer mag Schweizer?
a') Alyways this terrible traffic. After all, who likes Swiss people.
b) Nobody likes Swiss people => *topic*: Swiss people, also traffic
c) Prove by sentiment analysis, that nobody likes Swiss people (in traffic)

The example given here is already quite complex. First, it already suggests, that there is a negative sentiment in the first sentence, however that is determined.[20] The semantic information as well as the sentiment could be taken as first hint towards the interpretation of the question in the second sentence. Now there are two ways to continue. First, the simple approach that was suggested above, which would mean to make a sentiment analysis in Tweets dealing with Swiss people and base the judgment of the question on that.

For the second, deeper approach, it would be necessary for the machine to semantically connect Swiss people to traffic: *Swiss people cause terrible traffic.* The machine could then try to prove this hypothesis by analyzing tweets that contain both topics, Swiss people and traffic on their sentiment. Though, it is assumed here, that the first approach may be sufficient, the second, approach would be capable of describing the context of the RQ much better, or to say it differently, the reasoning would be more precise and logical.

Obviously, there is also another problem with the example given above. The topic established by the Tweet above, would mostly be understood in a certain geographic area, namely in Constance, since the traffic there is often terrible, especially on weekends. The point here is however, that for this approach to work properly, the context also needs to be restricted geographically. There could be other parameters as well, as for example hash-tags, since they already embrace Tweets of a certain topic. This means, if a Tweet is posted under a certain hash-tag, the context analysis could be used for Tweets under the same hash-tag. Both possibilities to extract tweets with a certain context, geographically and related to hash-tags, are executable with help of the Twitter API. However, it is not possible to analyze old Tweets, as already explained before.[21]

20 The easiest way would probably be to base the sentiment on the use of the adjective “furchtbar - awful/terrible”. This could be further refined by connecting it to always, which would imply, that there is a persistent problem with traffic.
21 https://dev.twitter.com/docs/api/1/get/search

All in all, this last approach may be the most efficient, when it comes to RQ-analysis on Twitter. However, its feasibility may be a problem. Although, the single steps presented here, do not pose to much of a challenge, the requirements posed by the Twitter API would make this kind of analysis only possible in a steady real-time analysis of tweets, since the tweets that would need to be analyzed as context, may change with every tweet. Without any proper data backing it up, it is assumed here, that the computing time would be fairly high. Although, it was said, that this may be the most efficient approach, it would require a link to a grammar-based approach, as well as a semantical approach. The necessity for a grammar explains itself, since the grammar needs to extract pieces of information, that are further processable by the semantic and sentiment analysis. The semantic analysis could be used in connection with the sentiment analysis. One reason is described as example below in (21).

(21) a) Wer hat denn noch eine Schreibmaschine?

It would be highly likely, that, if any sentiment analysis would be conductible at all, its result would be rather positive, for the nostalgic value of a type writer. Subtleties of that kind, might disturb the efficiency of the approach presented above. In this case, a semantic analysis based on the proposition, that someone has a typewriter could be executed. However, here a problem similar to the one for sentiment analysis could be encountered, because of the fact, that most people that do not have a type writer wouldn't bother to write about it at all. Before the thesis gets even more lost, the present section is summed up here. It should be clear, that the thought experiment proposed above, could go on endlessly. In conclusion the sentiment-analysis approach might be interesting. It is suggested however, that it should only be used in confined contexts, as proposed above. (hash-tags, geographic tags, tweets during a certain event etc.) Furthermore, it was shown that this approach would still require grammatical and semantic analysis, to derive the information, that needs to be processed by the sentiment analysis and search of contextual tweets. At last, it is assumed, that this approach also requires a large enough amount of tweets, about a certain topic to return adequate results.

3.3. Conclusion

All in all, the task to extract RQs is undeniably extremely complex. Some approaches were presented here, but as their discussion showed, they were mostly partial solutions and would probably need to be connected in some way, to get a feasible system. This insight is also supported by the theoretical part of this work. It was shown, that the method to produce a rhetorical question can vary depending on context and language. The problems that have been worked out here are even more prevalent in the English language, since it foremost lacks these certain particles, that help detecting RQs in German. In conclusion, automated processing of RQs in Twitter would require a large amount of work and probably would lead to explorations in artificial intelligence. However, it was shown, that theoretically all tools should be available. What needs to be done, is to assemble them in the right manner.

IV. Literature

Theoretical Part

Caponigro, Ivano & Jon Sprouse. 2007. Rhetorical Questions as Questions. *Proceedings of Sinn und Bedeutung 11*. p. 121-133

d'Avis, Franz Josef. 2001. Über >w-Exklamativsätze im Deutschen. Tübingen. Niemeyer

Grésillon, Almuth. 1980. Zum linguistischen Status rhetorischer Fragen. *Aufsätze.* Walter de Gruyter & Co

Grice, H.P. 1975. Logic and conversation. *Speech acts, volume III of Syntax and semantics.* P. 41-58.

Guerzoni, E.: 2003, Why 'Even' Ask? On the Pragmatics of Questions and the Semantics of Answers. PhD dissertation, MIT.

Guerzoni, Elena. 2004. EVEN-NPIs in Yes/No questions. *Natural Language Semantics 12.* p. 319-343

Gunglogson, Christine. 2002. Declarative Questions. *SALT XII.* p 124-143

Han, Chung-hye. 2002. Interpreting interrogatives as rhetorical questions. *Lingua 112*. p. 112–229

Koshik, Irene. 2002. A conversation analytic study of yes/no questions which convey reversed polarity assertions. *Journal of Pragmatics 34*. p. 1851-1877

Maibauer, Jörg. 1986. Rhetorische Fragen. Tübingen. Max Niemeyer Verlag

Rooy, Robert van. 2003. Negative Polarity Items in Questions: Strength and Relevance. *Journal of Semantics 20*. p. 239-273

Sadock, Jerry. 1971. Queclaratives. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society.*

Sadock, Jerry. 1974. Towards a Linguistic Theory of Speech Acts. Chicago. Academic Press

Stalnaker, Robert. 1978. Assertion. *Peter Cole (Hg.) Syntax and semantics 9: Pragmatics*. New York: Academic Press.

Truckenbrodt, Hubert. 2004. Zur Strukturbedeutung von Interrogativsätzen. *Beiträge aus Forschung und Anwendung*

Practical Part

Butt, Miriam et al. 1998. A Grammar Writer's Cookbook. Stanford, CA: CSLI Publications

Dent, K.; Paul, S. Through the Twitter glass. Workshop on Analyzing Microtext at the 25th AAAI Conference on Artificial Intelligence; 2011 August 8; San Francisco, CA.

Jurafsky, Daniel & Martin, James H. 2009. Speech and Language Processing. New Jersey. Pearson Education

Li, Baichuan et al. 2011. Question Identification on Twitter. CIKM'11. Glasgow, Scotland, UK.

Lichte, Timm. 2005. Corpus-based Acquisition of Complex Negative Polarity Items. *Proceedings of the Tenth ESSLLI Student Session Chapter 14*. p 157 – 168.

Ousterhout, John K. 1998. Scripting: Higher-Level Programming for the 21st Century. *IEEE Computer magazine march 1998*. p. 23-30

Paul, S.A., L. H., and Chi, E. 2011b. What is a Question? Crowdsourcing Tweet Categorization. Position paper at the *workshop on Crowdsourcing and Human Computation at the ACM Conference on Human Factors in Computing Systems* (CHI ’11).

Rosa, K., and Ellen, J. 2009. Text classification methodologies applied to micro-text in military chat. *In Machine Learning and Applications, 2009*. ICMLA ’09. International Conference. p. 710 –714.

Zhao, Zhe, Mei, Qiaozhu. 2013. Questions about questions: An Empirical Analysis of Information Needs on Twitter. International World Wide Web Conference Committee

Weblinks:
http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html (last accessed: 18.08.2014)

http://www.nltk.org/_modules/nltk/tokenize/punkt.html(last accessed: 09.08.2014)

http://www.nltk.org/(last accessed: 17.06.2014)

https://docs.python.org/2/tutorial/index.html(last accessed: 18.08.2014)

https://dev.twitter.com/docs/api/1/get/search(last accessed 15.08.2014)

V. Appendix

1. Data from the long night of sciences

a) Sample survey

b) Average rating of the test-items

c) collected definitions and detection cues for RQs

2. disk

a) RQ Analyzer (classes, readme, testsuite.txt)

Run Analyzer_main.py with python 2.7x in terminal.

Check readme for requirements (3rd party tools, path adjustments)

Thema:

Dieser Fragebogen dient als Studie zur Identifikation von rhetorischen Fragen ohne, bzw. mit nur sehr wenig Kontext. Die hier aufgeführten Fragen sind beispielhaft für Posts auf Social Media Seiten, wie zum Beispiel Twitter. Insbesondere Posts in Twitter unterliegen kontextuellen Restriktionen. Ein wichtiger Faktor dabei ist die Zeichenbeschränkung, sowie die Tatsache, dass private Gespräche anhand von Tweets kaum möglich sind. Alles wird öffentlich gepostet.

Durch die Analyse von Twitterdaten soll es möglich werden zu erfahren, welche Grundhaltung die Nutzer zu gewissen Themen, oder im Allgemeinen haben. Untersuchungen dieser Art könnten stark von der Analyse rhetorischer Fragen profitieren und somit bieten rhetorische Fragen ein interessantes Forschungsfeld. Dieser Fragebogen soll helfen, rhetorische Fragen besser zu verstehen. Dadurch soll es möglich werden, diese auch automatisch, also mit Hilfe von Computern, zu erkennen und zu bewerten.

Aufgabe:

Im Folgenden bitten wir darum, dass Sie die Fragen die sie nun lesen werden mit Rücksicht auf den Hintergrund(Twitter, Social Media) bewerten. Versuchen Sie möglichst genau zu bewerten, für wie rhetorisch Sie eine Frage halten. Überlegen Sie sich, ob manche Fragen rhetorischer wirken, als andere und lassen Sie dies in ihre Bewertung einfließen. Zur Bewertung steht ihnen eine Skala von 1 bis 10 zur Verfügung, wobei 1 *überhaupt nicht rhetorisch* und 10 *eindeutig rhetorisch bedeutet*. Im Anschluss würden wir Sie zusätzlich darum bitten, noch ein mal darüber nachzudenken, wie Sie die Fragen bewertet haben und aufgrunddessen eine kurze Definition für rhetorische Fragen abzugeben.

Erster Teil:

	Frage	**Bewertung** **1 = überhaupt nicht rhetorisch** (verlangt eine Antwort/Information) **10 = eindeutig rhetorisch**									
		1	2	3	4	5	6	7	8	9	10
1	Wer räumt denn schon sein Zimmer auf?	O	O	O	O	O	O	O	O	O	O
2	Wieso muss man eigentlich Gemüse essen?	O	O	O	O	O	O	O	O	O	O
3	Warum zum Teufel habe ich nicht genug gelernt?	O	O	O	O	O	O	O	O	O	O
4	Das ist ja nicht auszuhalten. Könnt Ihr nicht einfach alle mal ruhig sein?	O	O	O	O	O	O	O	O	O	O
5	Hast du etwa schon wieder den Müll nicht herausgebracht?	O	O	O	O	O	O	O	O	O	O
6	Weißt du, ob morgen Unterricht ist?	O	O	O	O	O	O	O	O	O	O
7	Warum ist die Welt so gemein zu mir?	O	O	O	O	O	O	O	O	O	O
8	Was habe ich denn nur verbrochen?	O	O	O	O	O	O	O	O	O	O
9	Wollt ihr mich alle auf den Arm nehmen?	O	O	O	O	O	O	O	O	O	O
10	Und das sollen wir den Politikern glauben?	O	O	O	O	O	O	O	O	O	O
11	Ich gehe jetzt ein Eis essen, wer kommt mit?	O	O	O	O	O	O	O	O	O	O
12	Wieso in Gottes Namen vergesse ich immer, wo ich meinen Schlüssel abgelegt habe?	O	O	O	O	O	O	O	O	O	O

13	Hat schon mal jemand diesen Film gesehen???	O	O	O	O	O	O	O	O	O	O
14	Weißt du, was du da tust??!	O	O	O	O	O	O	O	O	O	O
15	Fühlt ihr euch auch immer als hättet ihr gesündigt, nachdem ihr ein riesiges Stück Torte gegessen habt?	O	O	O	O	O	O	O	O	O	O
16	Wer hätte nicht mal Lust, nen Ferrari zu fahren?	O	O	O	O	O	O	O	O	O	O
17	Hat denn keiner eine Antwort darauf?!	O	O	O	O	O	O	O	O	O	O
18	Ihr wollt doch nicht wirklich alle jetzt schon nach Hause gehen?	O	O	O	O	O	O	O	O	O	O
19	Meint ihr nicht, dass ich das nicht auch gewusst habe?	O	O	O	O	O	O	O	O	O	O
20	Du weißt doch wie wir hier wieder herauskommen, nicht wahr?!	O	O	O	O	O	O	O	O	O	O
21	Hatte ich den Termin morgen oder übermorgen?	O	O	O	O	O	O	O	O	O	O
22	Musst du eigentlich immer alles besser wissen?	O	O	O	O	O	O	O	O	O	O
23	Habe ich dich jemals im Stich gelassen?	O	O	O	O	O	O	O	O	O	O
24	Wer hat denn da nicht aufgegessen?	O	O	O	O	O	O	O	O	O	O
25	Will ich wirklich wissen, woher diese Schmerzen kommen?	O	O	O	O	O	O	O	O	O	O
26	Was zur Hölle soll ich denn damit anfangen?	O	O	O	O	O	O	O	O	O	O
27	Du begleitest mich doch, ne?	O	O	O	O	O	O	O	O	O	O
28	Weiß irgendjemand wo wir gerade sind???	O	O	O	O	O	O	O	O	O	O
29	Wenn Gott gewollt hätte, dass auch ich einen Mann bekomme, warum hat er mich dann schlau und hübsch gemacht?	O	O	O	O	O	O	O	O	O	O
30	Es gibt doch noch Nachtisch, oder?	O	O	O	O	O	O	O	O	O	O
31	Wenn du es nicht warst, wer denn dann??!	O	O	O	O	O	O	O	O	O	O
32	Hat nicht jeder von uns schon mal nicht weiter gewusst?	O	O	O	O	O	O	O	O	O	O
33	Könnte mir mal bitte jemand erklären, warum es jetzt regnen muss?	O	O	O	O	O	O	O	O	O	O
34	Was könnte ich mir auch davon kaufen?	O	O	O	O	O	O	O	O	O	O
35	Und wie ist das Spiel gestern ausgegangen? Mal wieder verloren?	O	O	O	O	O	O	O	O	O	O
36	Mein Zuhause ist doch nur 10km entfernt. Warum in Gottes Namen muss ich hier schlafen?	O	O	O	O	O	O	O	O	O	O
37	Keiner von euch war gestern auf der Party???	O	O	O	O	O	O	O	O	O	O
38	Dein Anschluss ist abhörsicher? Mit wem telefonierst du denn? Al-Qaida?	O	O	O	O	O	O	O	O	O	O
39	Wo bekomme ich so kurzfristig noch ein Auto her?	O	O	O	O	O	O	O	O	O	O
40	Ehrlich jetzt? Habt ihr nicht mehr zu bieten?	O	O	O	O	O	O	O	O	O	O

41	Wer klingelt denn hier schon in aller Herrgottsfrühe???	O	O	O	O	O	O	O	O	O	O
42	Was hast du für Wünsche??!	O	O	O	O	O	O	O	O	O	O
43	Wer kann das schon wissen?	O	O	O	O	O	O	O	O	O	O
44	Wer hier ständig um Ansehen bettelt, hat doch kein stolz mehr, oder?	O	O	O	O	O	O	O	O	O	O
45	Sie ist nicht auf den Mund gefallen, hm?	O	O	O	O	O	O	O	O	O	O
46	Ob er wohl seine Hausaufgaben gemacht hat?	O	O	O	O	O	O	O	O	O	O
47	Ist er nicht putzig?	O	O	O	O	O	O	O	O	O	O
48	Wer putzt eigentlich immer das Klo??	O	O	O	O	O	O	O	O	O	O
49	Ist das denn zu glauben?	O	O	O	O	O	O	O	O	O	O
50	Wer braucht das schon zu lesen??	O	O	O	O	O	O	O	O	O	O

Zweiter Teil:

Bitte nehmen Sie sich nun noch ein mal einen Moment Zeit darüber nachzudenken, wie Sie die Fragen im ersten Teil des Bogens bewertet haben. Versuchen Sie wenigstens zwei Kriterien zu finden, an denen sie die Rhetorizität einer Frage festmachen würden, oder geben sie eine Definition an, anhand derer sie die Sätze identifiziert haben.

_____

_____

_____

_____

_____

_____

Average Ratings of the test-items

Weißt du, ob morgen Unterricht ist?	1.23
Ich gehe jetzt ein Eis essen, wer kommt mit?	1.73
Hatte ich den Termin morgen oder übermorgen?	1.98
Wo bekomme ich so kurzfristig noch ein Auto her?	2.41
Hat schon mal jemand diesen Film gesehen???	2.8
Weiß irgendjemand wo wir gerade sind???	2.98
Es gibt doch noch Nachtisch, oder?	3.54
Und wie ist das Spiel gestern ausgegangen? Mal wieder verloren?	3.8
Wieso muss man eigentlich Gemüse essen?	3.83
Wer putzt eigentlich immer das Klo??	4.24
Du weißt doch wie wir hier wieder herauskommen, nicht wahr?!	4.29
Ob er wohl seine Hausaufgaben gemacht hat?	4.49
Was hast du für Wünsche??!	4.63
Wenn du es nicht warst, wer denn dann??!	4.66
Fühlt ihr euch auch immer als hättet ihr gesündigt, nachdem ihr ein riesiges Stück Torte gegessen habt?	4.85
Keiner von euch war gestern auf der Party???	5
Wer klingelt denn hier schon in aller Herrgottsfrühe???	5
Hast du etwa schon wieder den Müll nicht herausgebracht?	5.39
Hat denn keiner eine Antwort darauf?!	5.56
Du begleitest mich doch, ne?	5.76
Das ist ja nicht auszuhalten. Könnt Ihr nicht einfach alle mal ruhig sein?	6.2
Weißt du, was du da tust??!	6.27
Mein Zuhause ist doch nur 10km entfernt. Warum in Gottes Namen muss ich hier schlafen?	6.34
Ihr wollt doch nicht wirklich alle jetzt schon nach Hause gehen?	6.37
Was könnte ich mir auch davon kaufen?	6.55
Meint ihr nicht, dass ich das nicht auch gewusst habe?	6.66

Was zur Hölle soll ich denn damit anfangen?	6.68
Ehrlich jetzt? Habt ihr nicht mehr zu bieten?	6.88
Sie ist nicht auf den Mund gefallen, hm?	7.07
Habe ich dich jemals im Stich gelassen?	7.1
Wieso in Gottes Namen vergesse ich immer, wo ich meinen Schlüssel abgelegt habe?	7.15
Wer hat denn da nicht aufgegessen?	7.29
Was habe ich denn nur verbrochen?	7.39
Warum zum Teufel habe ich nicht genug gelernt?	7.41
Wer hätte nicht mal Lust, nen Ferrari zu fahren?	7.41
Und das sollen wir den Politikern glauben?	7.51
Wer hier ständig um Ansehen bettelt, hat doch kein stolz mehr, oder?	7.59
Will ich wirklich wissen, woher diese Schmerzen kommen?	7.63
Wollt ihr mich alle auf den Arm nehmen?	7.65
Ist er nicht putzig?	7.68
Warum ist die Welt so gemein zu mir?	7.88
Wenn Gott gewollt hätte, dass auch ich einen Mann bekomme, warum hat er mich dann schlau und hübsch gemacht?	7.95
Hat nicht jeder von uns schon mal nicht weiter gewusst?	8.02
Musst du eigentlich immer alles besser wissen?	8.05
Dein Anschluss ist abhörsicher? Mit wem telefonierst du denn? Al-Qaida?	8.1
Wer braucht das schon zu lesen??	8.1
Wer kann das schon wissen?	8.32
Ist das denn zu glauben?	8.32
Wer räumt denn schon sein Zimmer auf?	8.66
Könnte mir mal bitte jemand erklären, warum es jetzt regnen muss?	8.76

Collected definitions and detection cues for RQs

-keine Antwort wird erwartet, weil die schon offensichtlich ist

-Dinge über die allgemeiner Konsens herrscht, werden ins Gegenteil verkehrt (teilweise spöttisch)

-Fragen die man bereits selbst beantworten kann

-Fragen die meine Meinung bestärken und bestätigen

rhetorische Fragen sind auf keine Antwort aus

-beziehen sich auf keinen konkreten Sachverhalt, keine spezifische Sache

-vorgestellte Intonation

-Anzahl der Fragezeichen

-rhetorische Fragen sind die, die einfach nur ein Statement machen

an zusätzlichen Wörtern wie “schon”

durch Andeutung mehrerer ? Welche den Tonfall andeuten

wenn es keine konkret angesprochenen Personen gibt

Satzzeichen-Intensität

Fragen zu erklärtem/logischem Sachverhalt

Bei “oder,hm ..” am Ende der Frage fand ich die Frage meist rhetorisch

Satzzeichen, Bekanntheitsgrad (Gebrauch), Überlegung ob man darauf antworten könnte

-keine Antwort wird verlangt

die Frageabsicht ist nicht Informationsgewinn

drückt das Gegenteil von dem aus was gesagt wird

wenn Antworten im vorhinein klar sind

wiederholte/mehrfache Satzzeichen

bestimmte Frageworte (oder,hm.ne)

Suggestionsfragen

gewisse Worte (denn, schon, eigentlich)

Flüche (zum Teufel in Gottes Names)

-Sinn der Frage

-keine Antwort wird erwartet

-manche sind keine Fragen sondern Ausrufe

-rhetorisch = verlang nicht unbedingt eine Antwort, d.h. Es ist wohl eher eine Feststellung

Anhand der Fragestellung ob eine Antwort gewünscht und realistisch ist und ob ein Sachverhalt übertrieben oder im Pathos geschildert wurde

Fragen, bei denen man die Antwort schon kennt, sich eine andere wünscht Fragen die an niemanden persönlich oder sich selbst gestellt werden

Übertreibung

viele Frage und Ausrufezeichen: eher rhetorisch

einige Fragen nach der Ursache eines Ereignisses, können auch informativen Charakter haben, können aber auch rhetorisch sein

Je leichter die Frage sachlich zu beantworten ist, desto weniger Rhetorizität weist sie auf

Bewertung entstammt aus Alltagssituationen

Bringt eine Antwort ein Ergebnis? Eine Erkenntnis?

Kann die Frage als solche stehen bleiben?

Es wird keine sinnvolle Antwort erwartet

der Satz ist Aussage an sich

Antwort wird halt erwartet bzw ist bereits klar

mit der Frage wird das Gegenteil ausgedrückt

1. Ist da Betoung durch den Fragenden

2. Lebenserfahrung, dadurch wird der Sinn der Fragen verstanden und Einordnung erfolgt – ob Frage

“ernst” gemeint ist, oder ob es sich um eine Frage handelt, die keine Antwort verlangt.

Wenn keine Antwort darauf gegeben werden kann

Es sind hauptsächlich Partikel die die rhetorische Frage identifizieren; plus: die doppelten ? Und !

Uneindeutig sind Fälle, weil sie keine Indikatoren haben und pragmatische Informationen fehlen

Antwort vorher schon bekannt

Satzteile wie “in Gottes Namen”, “zum Teufel”, etc

Rqs hatten eine stärkere Betonung auf die Aussage

Fragen die nicht intuitiv beantwortet werden wollten, auch bei Einschätzungsfragen

Eine rhetorische Frage liegt vor, wenn der/die Sprecher(in) beim Fragen nicht nach einer Information sucht. Der Wahrheitswert ist ihm/ihr schon bekannt, er/sie will nur eine gewisse Haltung in Bezug auf diesen Wert

echter Frageinhalt

konkrete Antwort möglich, die ein Handeln zu Folge hat

Eine rhetorische Frage verlangt nicht eindeutig nach einer Antwort, weil sie nicht immer direkt an eine Person gerichtet ist. (z.B. Warum regnet es heute?)

Wenn sie an eine Person gerichtet ist, kann die Antwort vielfältig sein

Die Fragen beiben oft ohne Antwort

Eine Frage wirkt um so Rhetorischer je surrealer der Inhalt der im Satz verwoben ist. Die Surrealität bezieht sich auf den Ort der Frage und der Verwendung von Superlativen Verben

Fragen die eine Antwort suggerieren (Manipulation)

Fragen die eigene Meinung bestätigen

Fragen die ohne eine direkte Antwort eine Information liefern

die Bemerkung soll humoristisch sein und man wäre zu höflich um eine ausgiebige u. ehrliche Antwort zu geben

um dem Sprecher einfach nur recht zu geben

es bedarf keiner weiteren Ausführung

Es wird tatsächlich keine Antwort erwartet, weil eine einvernehmliche/eindeutige Antwort

vorrausgesetzt wird (ähnliches einverständnis das auch bei Ironie vorausgesetzt wird)

etwas mehr Spielraum bei den Antwortmöglichkeiten

Eine zuvor unbekannte Antwort wird […] erfragt.

Sprecher hat eine gewisses Vorurteil im Bezug auf die erfragte Antwort

rhetorische Frage

Partikel denn

Wunschsatz 16

inhaltlich erkenntl. Dass nicht nach Information gesucht wird

ist eher eine Herausforderung als eine Frage

eine Antwort ist nicht erforderlich/nicht wichtig

es steckt eine Aufforderung in der Frage, die man nicht direkt formulieren will

Die Frage lässt Interpretationsspielraum für den Zuhörer

Sie erfordert eine Nachfrage

Die Semantik stand am Anfang im Fokus

Auf dem zweiten Blatt fielen mir die doppelten und dreifachen Satzzeichen auf

Einerseits am Inhalt, gibt es nur unmögliche Antworten darauf, ist die Frage rhetorisch

Fragen mit vielen Ausrufezeichen habe ich weniger oft als rhetorische Fragen gekennzeichnet

Satzzeichen

Intonation

Konjunktiv

Partikel

Häufigkeit der Sätze im Sprachgebrauch/bestimmte Floskeln

überlegt ob man auf die Frage eine Antwort hören möchte, oder nicht

ob die Fragen ironisch formuliert sind

keine Antwort wird erwartet

Gegenteil ist als Aussage gemeint

Fragen über Dinge die sowieso klar sind

oder-Ende

ist die Antwort auf die Frage unsinnig/übertrieben?

Ist die Frage emotional gestellt (??!!)

Jede Frage verlangt nach einer Reaktion des Angesprochenenen. Im Gegensatz zu “echten” Fragen zielen rhetorische jedoch nicht auf einen Informationsgewinn ab, sondern verlangen nach einer persönlichen Stellungnahme bzw. Rechtfertigung, nach Bestätigung, oder Anteilnahme oder einem sich Befassen mit dem durch die Frage angesprochenen Thema.

Die Rhetorizität wird dabei dadurch erkenntlich, dass die Frage zusätzliche Informationen enthält, die nicht zur Kennzeichnung einer benötigten Information alleine nötig wären. Dies können emotionale Färbungen bzw angedeutete Wertungen des Fragers (schon) sein, Erwartungshaltungen (‚ne) etc

Im Extremfall ist die Frage zwar rein formal auf Informationserwerb gerichtet, aber inhaltlich nicht zu beantworten, Hier steckt dann die Information welce auf die Rhetorizität deutet in der Natur des Inhalts der Frage, wo Frager und Befragter wissen, dass der jeweils andere sich der offensichtlichen Nichtbeantwortbarkeit im Klaren Ist und dieses Wissen ebenfalls auch beim Gegenüber vorraussetzt. (Wissen über gemeinsames Wissen)

Bei rhetorischen Fragen sind dem Fragesteller die Antworten schon klar, bzw lässt er nur eine Antwort als Richtige gelten und interessiert sich nicht wirklich für die Meinung seines gegenübers.