potwag:rs
Potsdam Tweet Annotation Guidelines: Rhetorical Structure

Wladimir Sidorenko, Manfred Stede

March 24, 2015

Contents

TODO

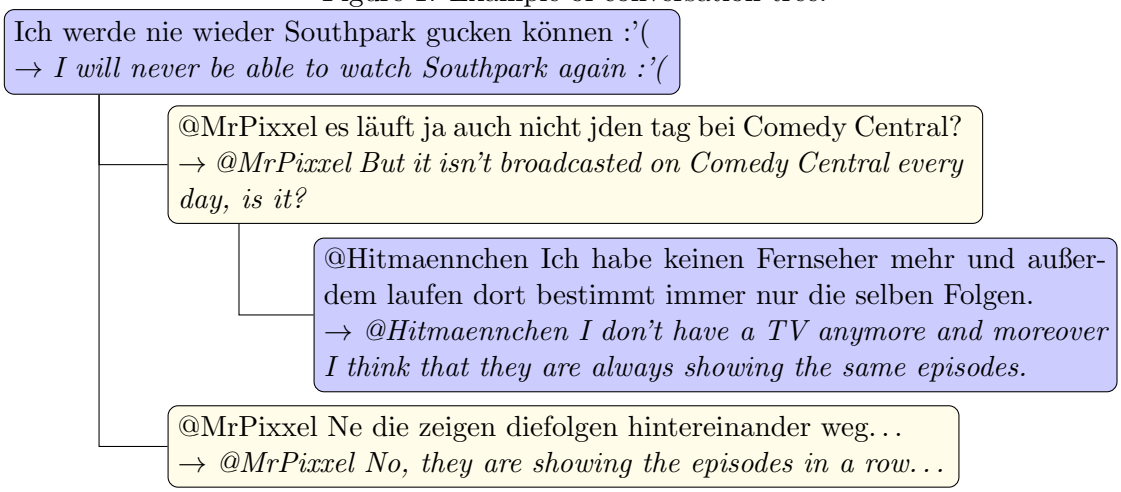
- alle Rel-Definitionen ueberpruefen. Dabei Justify and Reason besser voneinander abgrenzen, dito JOINT und LIST. Diese beiden mssen auch von SEQ abgegrenzt werden - SEQ hat Vorrang, wenn temporale Folge (aufsteigend oder absteigend!) von Ereignissen gegeben ist; wenn nicht, prfe JOINT/LIST.
- Jede Relation sollte 1-2 maz-Beispiele haben, um ihren Gehalt mglichst gut zu illustrieren, plus 1-2 tweet Beispiele fr die Praxis.
- Beim Topic-Wechsel verwenden wir die Relation Joint: Beispiel: die Energiewende ist gerecht genug! – Joint – Redet doch mal vom #klimaschutz @SZ — @MCSvDrach — so ein quatsch, — die Energiewende ist gerecht genug! — Kaum jemand wechselt den Stromanbieter! — Redet doch mal vom #klimaschutz
- Rhetorische Fragen behandeln
- je mehr authentische twitter Beispiele desto besser.
- INTER-TWEET: Prinzip ist, dass jeder tweet einen wohlgeformten RST Baum bekommt und ber eine Relation als Satellit an den Vorgngertweet angebunden wird. ZUSTZLICH kann beim Annotieren entschieden werden, dass nur ein Segment des Tweets in einer bestimmten Relation zum Vorgngertweet oder einem Teil davon steht. Dann kann eine "sekundre" Relation etabliert werden. Wir fordern dann, dass **alle** EDUs (oder bergeordnete Segmente) eine solche sekundre Anbindung erhalten. Fr die Ziele der Anbindung gibt es keine Vorgaben (beliebige Segmente)

1 Introduction

These guidelines present instructions for the annotation of elementary discourse units (EDUs) and rhetorical structure relations on Twitter. For the time being, the target language of the annotation is German, but an adaptation to English should not be difficult. Largely based on the classical works on Rhetorical Structure Theory (RST) (Mann

and Thompson, 1988; Carlson et al., 2001; Taboada and Mann, 2006), these instructions should be seen as an extension and adaptation of RST to the peculiarities of microblog conversations. Such conversations arise when one tweet directly replies to another one, which yields a data structure of a tree rooted in the initial, topic-initiating tweet (cf. Figure 1).

Figure 1: Example of conversation tree.



Our proposed extensions concern both the definition of EDUs and the set of rhetorical relations that are used for them. For the former, we extend the traditional notion of discourse units to account for such microblog-specific phenomena as emoticons, @-mentions, URIs, and #hashtags. For the latter, we, on the one hand, restrict the originally proposed set of discourse relations in order to reduce their ambiguity. On the other hand, we introduce a few additional Twitter-specific relation types and also add a completely new group of so-called inter-message relations which are discourse connections that hold between two or more different microblogs.

The basic approach would be to assign an RST structure to each individual tweet (much like in ordinary text analysis) and then to link this tweet's tree via an "inter-tweet" relation to the tree of its parent message (i.e., the message that the former tweet replies to). At any point in the annotation process, the annotator is shown the current tweet (henceforth: the *target*) plus the parent tweet (henceforth: the *reference*) that has already been annotated, along with the RST-tree for the latter. This pair is what we regard as the linguistic unit to which *coherence* can be reasonably ascribed: A user starting her reply will also often be aware only of the parent tweet and not of other parts of the conversation (although this might differ between the technical types of twitter clients).

Twitter conversations are presented to the annotator in an incremental fashion (reply by reply). When shown the reference tweet and the target, the annotator's task therefore would be *a*) to determine the boundaries of discourse units within the target, *b*) to construct the RST tree for the target tweet using the set of intra-tweet relations, and, finally, *c*) to determine the type of the inter-tweet relation between the target and its reference message and to draw a corresponding edge between their trees.

The rest of this document is organized as follows. The following subsection 1.1 discribes format of these guidelines. Section 2 provides a definition of elementary discourse units

and describes criteria for determining the boundaries of EDUs in tweets. Section 3 presents the rhetorical relations defined in this annotation schema and also gives some examples of their use. Section 4 provides a general description of the annotation process. Additional annotation hints and explanations of controversial questions can be found in Sections 3.6 and 5 respectively.

1.1 Format of the guidelines

In these guidelines, we rely on the following conventions regarding the formatting of text. We use the `typewriter font` for literal mentions of shell commands, the name of the annotation tool, mentions of its parts and buttons. Text examples and definitions are shown in *italics*, and English translations of German examples start with a →. The boundaries of EDU segments are demarcated with square brackets [...], and the corresponding numbers of the segments are given as a subscript after their right closing brackets: [*An example segment might look like this*]$_1$. For literal mentions of rhetorical relations, we use the SMALL CAPS FONT.

2 Segmentation

For spoken language or for social media contributions, defining the notion of elementary discourse unit (EDU) is not easy, and no consensus to this end has been achieved in the literature. For the present annotation task, we rely on formal, functional, and semantic criteria alike, and thus propose the following working definition:

Definition 1. ***Elementary discourse units*** *are non-overlapping spans of text that within their context express complete (but possibly elliptic) propositions or speech acts, are syntactically independent of their surrounding context, and serve a clearly distinguishable informative, pragmatic, or communicative purpose.*

According to this description, an elementary discourse unit should never overlap with any other EDUs. The omission of this unit should not break any syntactic or grammatical dependencies outside of this unit, but its omission should significantly change the amount of information conveyed by its encompassing message or crucially modify the pragmatic or communicative purpose of that message. Importantly, the "completeness" criterion is not meant as strict syntactic completeness: ellipses (of NPs or VPs) are widespread in our genre, and their presence does not preclude the existence of a complete *message*, provided that the reader can easily resolve the ellipses in their context.

A sample tweet that has been divided in elementary discourse segments is provided below:

(1) [*Wäre ich nicht so faul,*]$_1$ [*würde ich rüber zum McDonalds gehen.*]$_2$ [*Aber faul.*]$_3$
→[*If I wasn't so lazy,*]$_1$ [*I would go over to McDonalds.*]$_2$ [*But I'm lazy.*]$_3$

As you can see, none of the EDUs overlaps with any other unit (none of the words belongs to two EDUs simultaneously). The deletion of either segment does not break any

syntactic constituents ("*I would go over to McDonalds*" is syntactically valid, while "*I to McDonalds*" is not). This deletion, however, deprives the message of a significant pragmatic information ("*If I wasn't lazy, I would go over to McDonalds*" means that one actually is not going to visit the fast food restaurant, while "*I would go over to McDonalds*" denotes the opposite).

Elementary discourse units are typically represented by:

- clauses;
- certain prepositional phrases;
- parentheticals;
- interjections or interjection-like elements;
- Twitter-specific elements like @-mentions, hyperlinks, etc.

In the following, we provide a more detailed description and segmentation rules for each of these elements.

2.1 Clauses

We regard the following types of clauses as separate EDUs:

- main clauses, e.g. [*Ich muss bald zum Friseur*]$_1$ → [*I have to visit the hairdresser soon*]$_1$;
- coordinate main clauses, e.g. [*Kann man wohl sagen,*]$_1$ [*aber der Chef hört sowas nicht gerne*]$_2$ → [*One could say so*]$_1$ [*but the boss doesn't like hearing that*]$_2$;
- non-restrictive relative clauses, e.g. [*Ich habe eine neue Idee,*]$_1$ [*die die Welt verändern wird*]$_2$ → [*I have a new idea,*]$_1$ [*which will change the world*]$_2$[1]
- adverbial subordinate clauses, e.g. [*Ne 6 wäre es,*]$_1$ [*wenn ich schwänze*]$_2$ → [*It would be a 6,*]$_1$ [*if I cut the lesson*]$_2$

When adverbial subordinated clauses are coordinated or enumerated, then each of them forms a separate EDU:

(2) [*Wenn ich zum Friseur muss*]$_1$ [*oder neue Schuhe brauche,*]$_2$ [*fahre ich mit dem Auto nach Köln,*]$_3$ → [*When I need to see the hairdresser*]$_1$ [*or buy new shoes,*]$_2$ [*I go by car to Cologne*]$_3$

Importantly, complement clauses that specify the "content" of the governing verb do not form EDUs, because they are syntactically and semantically an integral part of the matrix clause. This rule also applies to clauses that constitute the subject or the object of the matrix clause. For example:

[1] A non-restrictive relative clause provides a separate message in which the referent of the NP head plays a role. In contrast, *restrictive* relative clauses merely serve to identify the referent, as in [*Wir entschieden uns für das Auto, das rechts am Rand stand*]$_1$ → [*We picked the car that stood at the right edge*]$_1$

(3) [*Ich denke dass ich bald zum Friseur muss*]$_1$ → [*I think that I have to visit the hairdresser soon*]$_1$

(4) [*Wer zum Friseur muss soll sich beeilen*]$_1$ → [*Those who have to visit the hairdresser shoud hurry up*]$_1$

(5) [*Ich verstehe was Du gesagt hast*]$_1$ → [*I understand what you said*]$_1$

A complication arises when such "content clauses" are coordinated: *I thought that the journey is long and the destination not worth the trouble.* Two (or, in other cases, more) separate thoughts are being expressed and should form distinct EDUs, whereby the second coordinate clause is to be interpreted as heavily elided:

(6) [*Ich fand dass die Reise lang ist*]$_1$ [*und den Aufwand nicht wert.*]$_2$ → [*I thought that the journey is long*]$_1$ [*and the destination not worth the trouble.*]$_2$

In general, coordination of clauses blends into coordination of smaller constituents (especially NPs), and ellipsis often makes the EDU decision difficult. Does *I bought an apple and an orange* denote two separate buying events and thus two distinct messages (to be paraphrased as *I bought an apple and I bought an orange*)? The borderline cannot be defined sharply. The annotator should opt for separating EDUs only when two different *kinds* of eventualities can be recognized. In the example just given, there is thus only one EDU. An indication for distinct eventualities is the presence of deverbal nouns. Notice that in general, the type of conjunction (additive, contrastive) should not make a difference for the decision. Examples:

(7) [*Ich heiße nicht Paul, sondern Peter*]$_1$ → [*My name is not Paul but Peter*]$_1$

(8) [*Ich kaufe kein Auto,*]$_1$ [*sondern miete eines*]$_2$ → [*I donn't buy a car*]$_1$ [*but I rent one*]$_2$

(9) [*Ich höre gern Klassik oder Jazz*]$_1$ → [*I like to listen to classical music or to jazz*]$_1$

TODO: DIES IST WOHL DER POTENZIELL KOMPLIZIETESTE TEIL... WIR MÜSSTEN AUCH NOCH ERWÄHNEN, WANN MAN KOORDINATIVE VPs ABTRENNEN SOLLTE, Z.B. "ICH KANN TANZEN UND SINGEN" oder "WENN ICH MIR HAARE SCHNEIDEN WILL, GEHE ICH ZUM FRISEUR ODER FAHRE IN DEN NÄCHSTEN SCHÖNHEITSSALON."

2.2 Prepositional phrases

Occasionally, we treat prepositional phrases as separate EDUs. This is restricted to cases where the embeded noun phrase denotes an eventuality (often by means of a deverbal noun) and there is a clearly-recognizable coherence relation to the host clause. Typical prepositions used in such a setting are *trotz* and *wegen* (despite, due to/because of).

(10) [*Ich muss bald zum Friseur*]$_1$ [*wegen der morgigen Theateraufführung*]$_2$ → [*I have to visit the hairdresser soon*]$_1$ [*because of tomorrow's theater play*]$_2$

2.3 Embedded segments and parentheticals

As stated in the beginning, EDUs cannot overlap: any word must belong to exactly one EDU. This does not preclude the embedding of EDUs, however. With relative or adverbial clauses, embedding happens quite often.

(11) [*Ich werde,* [*obwohl Benzin so teuer ist,*]$_2$ *wieder ein Auto kaufen.*]$_1$ → [*I will,* [*even though gas is so expensive,*]$_2$ *buy a car again.*]$_1$

However, the most frequent (and often more difficult) case of embedding is appositions and parentheticals: fragmentary material that is inserted inside an EDU. The basic criterion for deciding whether it forms a separate EDU is our standard rule: Is there a distinct message? Is there an eventuality being described? Again, ellipsis can complicate matters, especially when connectives are used that "usually" serve to conjoin EDUs. In the first example below, *Dichter* works like a restrictive relative clause, while in the second, the paranthesis provides a separate message about the speaker. In the third, the parenthetical provides an additional attribute, but it is quite similar to the first one and thus is not judged to be a separate message. In the final example, the attributes are quite distinct, and furthermore there is a clear coherence relation (Concession) present.

(12) [*Ich habe Goethe, den Dichter, nie gemocht.*]$_1$ → [*I never liked Goethe, the poet.*]$_1$

(13) [*Ich* [*(wie immer bescheiden)*]$_2$ *kaufe mir einen Porsche.*]$_1$ → [*I* [*(as usual being modest)*]$_2$ *buy a Porsche.*]$_1$

(14) [*Ein großer – und auch berhmter – Dichter sollte keine Lehrbücher schreiben.*]$_1$ → [*A great – and also famous – poet should not write textbooks.*]$_1$

(15) [*Ein erfolgloser* [*– wenn auch reicher –*]$_2$ *Dichter sollte keine Lehrbücher schreiben.*]$_1$ → [*An unsuccessful* [*– even if rich –*]$_2$ *poet should not write textbooks.*]$_1$

When parenthetical material appears at the end of a segment, we do not embed it but always treat it as a disjoint EDU.

2.4 Interjections and emoticons

As is well known, authors of tweets do not always follow the rules of well-formed written language. Interjections, as known in speech, are a class of phenomena that is related to the parentheticals we just described. We use "interjection" here in a wide sense, covering small portions of linguistic material that interrupts the syntactic unit. And in social media, *emoticons* are frequently inserted into the message, which we treat on a par with interjections, as they play essentially the same role.

Following our basic criterion of "messagehood", we use a deletion test to decide whether an interjection constitutes a separate EDU or not. If the deletion does not significantly alter the message, then the interjection is either relatively meaningless, or it merely reiterates a function that the message has already. This holds in the same way for emoticons: They

can, for example, change the message by conveying that is meant ironically, or merely add a smiling face to a message that is already recognizable as friendly. (This distinction, of course, can sometimes be difficult.) A general rule for emoticons is that they are never separate EDUs when part of a question.

(16) [*Das muss jetzt,* [*tutmirleid,*]$_2$ *echt schnell gehen.*]$_1$ (= Das muss jetzt echt schnell gehen; tut mir leid.) → [*This has to* [*I'm sorry,*]$_2$ *move really fast now.*]$_1$

(17) [*Das muss jetzt, ich mein, echt schnell gehen*]$_1$ (= Das muss jetzt echt schnell gehen) → [*This has to, I mean, move really fast now.*]$_1$

(18) [*Wer glaubt das schon? :-)*]$_1$ → [*Who believes that anyway? :-)*]$_1$

(19) [*Das ist eine interessante Geschichte*]$_1$ [*8-o*]$_2$ (= Das ist eine interessante Geschichte. Ich bin sehr überrascht.) → [*That's an interesting story*]$_1$ [*8-o*]$_2$ (= That's an interesting story. I'm astonished.)

Interjective material can also appear at the beginning or end of sentences. We use the presence of a full stop as an indicator for an EDU boundary; presumably, the author intended the following to be a separate message. As with parentheticals (and as shown in the emoticon example above), we do not embed material at the boundary of an EDU, but always introduce a separate EDU.

(20) [*Wir kommen pünktlich, logisch*]$_1$ → [*We'll be on time, clearly.*]$_1$

(21) [*Wir kommen pünktlich.*]$_1$ [*Logisch.*]$_2$ → [*We'll be on time.*]$_1$ [*That's clear.*]$_2$

Tag questions at the end of a tweet are separate EDUs when the annotator regards them as "real" instructions for the audience rather than politeness markers. That is, an *", oder?" / ", ja?" / "nicht?"* needs to be judged whether the author in fact intends to elicit a response by using them.

2.5 #, @, URI

Finally, three devices often found in microblogs are treated as separate EDUs when they are not integrated in the syntactic structure of the sentence: #hashtags, @-mentions, and URIs. Examples:

(22) [*riecht man es immer noch???*]$_1$ [*#Wittmund*]$_2$ [*#Kaverne*]$_3$
→ [*can you still smell it???*]$_1$ [*#Wittmund*]$_2$ [*#cavern*]$_2$
[*#Wittmund erinnert mich an die Frage, ob man es noch immer riecht.*]$_1$
[*#Kaverne*]$_2$
→ [*#Wittmund reminds me of the question whether one can still smell it.*]$_1$
[*#cavern*]$_2$

(23) [*@nrsss*]$_1$ [*dann darfst du dich auch nicht beklagen*]$_2$
→ [*@nrsss*]$_1$ [*then you should not complain*]$_2$
[*ich würde zu @nrsss sagen, dass er sich nicht beklagen soll*]$_1$
→ [*I would say to @nrsss that he should not complain*]$_1$

(24) [*Schau Dir das mal an:*]$_1$ [*http://px12.uvw*]$_2$
→ [*Check this out:*]$_1$ [*http://px12.uvw*]$_2$

3 Relations

For our annotation scenario, the notion of rhetorical relation in the sense of Mann and Thompson (1988) or the RST website[2] needs to be adapted in two ways: It needs to account for some phenomena that show up in dialogue (as opposed to monologue text), and it needs to cover phenomena that are specific to Twitter conversations.

In general, we retain the RST distinction between 'semantic' (or 'subject-matter') and 'pragmatic' (or 'presentational') relations (cf. Mann and Thompson (1988)). Relations of these two groups can hold both within one tweet and across two messages. In the former case, we call them *intra-tweet* and, in the latter case, we refer to them as *inter-tweet* relations.

In addition to this, we also introduce two more classes of discourse relations: *textual* and *communicative*. The former class consists of relations that serve solely text organizational purposes and do not convey any distinct semantic or pragmatic meaning. Relations of this type are purely *intra-tweet* (i.e. can only hold within one message). The second class, on the contrary, has a purely *inter-tweet* character and comprises relations which fulfill some communicative functions such as greeting, questioning, or answering a question.

In summary, the four basic groups can be characterized as follows:

- Primarily *semantic* relations are used when a complex state of affairs in the world is being described, for instance when it involves a causal relation between two events. The "point" of the relations is not to change the readers' attitudes but merely to inform them about the event(s).
- Primarily *pragmatic* relations concern the author's attempt to influence attitudes of the readers: motivate them to act, make them believe something, enable them to do something. Providing reasons for such actions is a central instrument involving a pragmatic relation between two segments.
- Primarily *textual* relations have a purpose of organizing the discourse. That is, they do not carry a clearly identifiable semantic or pragmatic load. A clear example is the List relation, which groups together various segments that collectively play the same role in the overall rhetorical structure.
- Primarily *communicative* relations serve the purpose of organizing the communication. This includes cases of managing the initiative (turn-taking) in the multilogue, asking for additional information, starting or concluding a discussion.

From RST we also inherit the distinction between *mononuclear* and *multinuclear* relations (which in principle may occur in any of the aforementioned groups; hence it is a cross-category distinction). Mononculear relations consist of a more important span – the

[2] http://www.sfu.ca/rst

nucleus – and a less important, supportive span – the *satellite*. A rule of thumb is: In a good 'summary' consisting of the most important segments of a text/tweet, the nuclear segments would be present, whereas the satellites would not.

When analyzing text, situations of relation *ambiguity* can arise fairly often, and many of them involve relations from the separate groups – i.e., there is both a semantic and a pragmagtic relation to be perceived. How to proceed in such a situation is discussed later in Section 3.6.

3.1 Format of the definitions

The definitions of relations characterize the roles of the connected segments and the effect that the author wants to achieve by making use of the relation. We provide linguistic examples, first from newspaper text (Potsdam Commentary Corpus) in order to more clearly convey the intended meaning of the relation, and then from authentic tweets. Further, we mention some typical connectives that can be used for the relation, and add comments on specific aspects for the relation occurring in tweets. The overall format of a relation definition is as follows:

- **N:** Description of type and/or function of the nucleus
  (characterization of the author's stance, not of specific linguistic expression)
- **S:** Description of type and/or function of the satellite
  (characterization of the author's stance, not of specific linguistic expression)
- **N/S:** Description of the function of the nuc/sat combination. If there are tendencies on linear order between nuc and sat, they are also mentioned here.
- **Effect:** Description of the effect intended by the author in using the relation; ("before/after")
- **Typical connectives**
- **Example:** In examples, N and S are marked. If a preceding context is given, it is set in *italics*.
- **Comment** (optional)

The entries for N, S and N/S remain empty if there is no such constraint for the relation. For multinuclear relations, the S and N/S entries are not given.

For the entry on possible connectives, notice that:

- We provide merely a few examples here; in general there will be more connectives that can mark the relation (often there are quite many)
- Often, connectives are ambiguous. On the one hand, a word can have a different reading in addition to the connective one (e.g., as a discourse particle). On the other hand, the same connective might be able to signal multiple relations, such as *since*, which can indicate a temporal relation or a CAUSE/JUSTIFY:

  - [*Since you first visited the city*]$_S$ [*many new buildings have been constructed.*]$_N$
  - [*Since you are smart*]$_S$ [*you will surely pass the exam.*]$_N$

- Occasionally, the relation that is strongly signalled by the connective is not the intended, "deep" relation. For example, a temporal relation may be stated, but there is causality "behind" it. In such a case, the "deep" relation should be annotated. A German example:
  - [*Kaum hatte sie ihre Rede begonnen,*]$_S$ [*waren alle Zuhörer begeistert.*]$_N$

In the annotation process, the applicability of a relation for linking two segments should be tested in this way:

- Is there a connecive in one of the segments, which signals the relation or at least restricts the set of possible relation candidates? (But note the caveat mentioned above – the connective might hide a deeper relation.)
- Punctuation is not a precise cue, but there are some tendencies for correlation, such as between semicolon and contrastive relations:
  - [*My brother likes swimming in lakes;*]$_S$ [*I never do that.*]$_N$
- By juxtaposing the two segments, does the author intend to achieve the effect given in the relation definition? This is the central and necessary condition for applying a relation.
- If the relation definition states constraints on the type or function of N or S or their combination, check whether these are fulfilled. These are also necessary conditions.
- Is the N/S distribution that is associated with applying the relation adequate? Does it capture the relative "weight" of the segments for the overall intention of the tweet? – This is a less strict criterion, but it can help in deciding between several relations that seem applicable.

In all definitions, we use W for 'writer', R for 'reader', N for 'nucleus', and S for 'satellite'.

3.2 "Semantic" intra-tweet relations

Circumstance

- **S:** A (non-hypothetical) state of affairs.
- **N/S:** S sets a framework in the subject matter within which R is intended to interpret N.
- **Effect:** R recognizes that S provides the framework for interpreting N.
- **Typical connectives:** *when; while; ...; wenn; während; ...* for temporal circumstantials

- **Example:** [*Als die Veag wegen der Liberalisierung des Strommarktes unter Druck geriet,*]$_S$ [*hielt sie sich bei ihren Lieferanten schadlos.*]$_N$ (maz-5297)

Condition

- **S:** A hypothetical, future, or otherwise unrealized situation (relative to the situational context of S).
- **N/S:** Realization of N depends on realization of S.
- **Effect:** R recognizes how the realization of N depends on the realization of S.
- **Typical connectives:** *if .. then; ... / wenn .. dann; sofern; ...*
- **Examples:** [*Wenn immer mehr Kommunen finanziell ausbluten,*]$_S$ [*wird die regelmäßige Schulbucherneuerung zur Illusion.*]$_N$ (maz-00002)
  [*Kommt der Sanitärtrakt nicht spätestens zur nächsten Saison,*]$_S$ [*droht Radewege den Wettbewerb um Freizeitkapitäne zu verlieren.*]$_N$ (maz-6488)
- **Comment:**

Otherwise

- **N:** A hypothetical, future or in some other way unrealized state of affairs.
- **S:** A hypothetical, future or in some other way unrealized state of affairs.
- **N/S:** The realization of N prevents realization of S.
- **Effect:** R recognizes the dependency relation of prevention between the realization of N and the realization of S.
- **Typical connectives:** *otherwise; ... / ehe; bevor; sonst; ...*
- **Examples:** [*Die Stadt wäre gut beraten, sich diesem moralischen Imperativ zu beugen.*]$_N$ [*Anderenfalls müsste sie vor jeder Schule und jeder Kita Bremsbuckel aufpflastern.*]$_S$ (maz-5709)
  [*Der Gemeinderat sollte den unseligen Beschluss aufheben,*]$_N$ [*ehe die Bürger sich mit Taschenlampen versorgen.*]$_S$ (maz-15609)

Unless

- **N:** A hypothetical, future or in some other way unrealized state of affairs.
- **S:** A hypothetical, future or in some other way unrealized state of affairs.
- **N/S:** S affects the realization of N; N is realized provided that S is not realized.

- **Effect:** R recognizes that N is realized provided that S is not realized.
- **Typical connectives:** *unless; ... / es sei denn; ...*
- **Example:** [*Morgen wird der Satellit in den Pazifik stürzen.*]$_N$ [*Es sei denn, er verglüht doch noch vollständig in der Erdatmosphäre.*]$_S$

Elaboration

- **N/S:** S presents additional detail about the situation (*not* about an individual entity – cf. E-Elaboration below) that is presented in N or inferentially accessible in N. Some typical relations between N and S are abstraction–instance, event–subevent.
- **Effect:** R recognizes S as providing additional detail for N.
- **Typical connectives:** *in particular; for example; ... / besonders; beispielsweise; ...*
- **Example:** [*Diepensee siedelt um.*]$_N$ [*Ohne Wenn und Aber.*]$_S$ (maz-6993)

E-Elaboration

- **N/S:** S presents additional detail about an entity mentioned in N. N precedes S in the text. Some typical relations between N and S are set–member and whole–part.
- **Effect:** R recognizes S as providing additional detail about an entity mentioned in N.
- **Typical connectives:** (rarely marked by connectives)
- **Example:** [*Heute wollen Tausende das Flair eines historischen Klassenzimmers kennen lernen,*]$_N$ [*wie sie nur noch wenige in Deutschland existieren.*]$_S$ (maz-6728)
- **Comment:** In the example, the *wie* minor clause (S) does not relate to the whole event of acquainting but only to the class room, i.e., an entity within N. .

Interpretation

- **N/S:** S relates N to a framework of ideas not involved in N itself and not concerned with W's positive regard; or at least, positive regard is not the important point. In the text, N precedes S.
- **Effect:** R recognizes the interpreting function of S for N.
- **Typical connectives:** *thus; ... / so; damit; mithin; ...*
- **Example:** [*In Potsdam findet heute der große Festzug zu Ehren von Friedrich II. statt.*]$_N$ [*Damit sind die diesjährigen "Preußen-Tage" beendet.*]$_S$
- **Comment:** If the main point of S is to assess N on a scale, the suitable relation is not Interpretation but Evaluation-N/Evaluation-S.

Means

- **N:** An activity.
- **N/S:** S presents a method or instrument which tends to make realization of N more likely.
- **Effect:** R recognizes that the method or instrument in S tends to make realization of N more likely.
- **Typical connectives:** *to this end; ... / dazu; damit; ...*
- **Examples:** [*Berliner fahren im August immer gern nach Lichtenrade.*]$_N$ [*Dazu nehmen sie meistens die S25.*]$_S$
  [*Der Parteivorsitzend hat gleich zwei CDU-Bürgermeister gegen sich aufgebracht*]$_S$ [*und wird damit Ansehen und Rückhalt verlieren.*]$_N$ (maz-18160)

Cause

- **N:** A state of affairs holding in the world. This state has been caused by the situation presented in S.
- **S:** A state of affairs holding in the world.
- **N/S:** The state of affairs in N is being caused by the state of affairs in S.
- **Effect:** R recognizes the causal relationship between the two states of affairs.
- **Typical connectives:** *because; since; ... / weil; da; deshalb; ...*
- **Example:** *Unerwartete Pressemeldung zur Verlegung des Ortes Diepensee.* [*Überrascht reagierte auch Bürgermeister Jochen Wagner.*]$_N$ [*Schließlich gaben die Stadtverordneten erst Montagabend grünes Licht für die weitere Erschließung des neuen Ortsteils Diepensee.*]$_S$ (maz-6993)

Result

- **N:** A state of affairs holding in the world. This state provides the reason for the situation presented in S.
- **S:** A state of affairs holding in the world.
- **N/S:** The state of affairs in S is being caused by the state of affairs in N.
- **Effect:** R recognizes the causal relationship between the two states of affairs.
- **Typical connectives:** *because; since; ... / weil; da; deshalb; ...*
- **Comment:** Dual of Cause. The choice between Cause and Result is made solely on the grounds of relative importance for the tweet.

Purpose

- **N:** An activity.
- **S:** A hypothetical, future or in some other way unrealized state of affairs. This state is expected to be achieved after the activity described in N is fulfilled.
- **N/S:** S is to be realized through the activity in N.
- **Effect:** R recognizes that the activity in N is initiated in order to realize S.
- **Typical connectives:** *by means of; in order to; ... / durch; um .. zu; ...*
- **Example:** [*Um ihre Soldaten zu schützen,*]$_S$ [*äußern sich die USA nicht zu Berichten über den Beginn der Boden-Operation in Afghanistan.*]$_N$ (maz-5701)
- **Comment:** There is, in a wide sense, a causal relation between S and N. The difference to the relations CAUSE/RESULT is that for PURPOSE, S is an unrealized state of affairs that is the intention of the actor.

Solutionhood

- **S:** S presents a problem.
- **N/S:** N is a solution to the problem presented in S. In the text, S usually precedes N.
- **Effect:** R recognizes N as a solution to the problem presented in S
- **Typical connectives:** (rarely marked by connectives)
- **Example:** [*Mit der Verabschiedung des Nichtraucherschutzgesetzes sitzen viele Kneipen in der Falle.*]$_S$ [*Es empfiehlt sich, früh genug auf die Einrichtung abtrennbarer Räume zu achten.*]$_N$

Contrast (multinuclear)

- **N:** No more than two nuclei; the situations in these two nuclei are (a) comprehended as the same in many respects (b) comprehended as differing in a few respects and (c) compared with respect to one or more of these differences.
- **Effect:** R recognizes the comparability and the difference(s) yielded by the comparison is being made.
- **Typical connectives:** *on the other hand; but; while; ... / demgegenüber; hingegen; während; aber; ...*
- **Example:** [*Mein erstes Auto war ein Kleinwagen.*]$_N$ [*Das zweite hingegen ein ausgewachsener Kombi.*]$_N$

Sequence (multinuclear)

- **N:** The nuclei describe states of affairs in the world that take place in a temporal order.
- **Effect:** R recognizes the temporal succession relationships among the nuclei..
- **Typical connectives:** *then; afterwards; before; ... / dann; anschließend; und; zuvor, ...*
- **Example:** [*Um neun betrat die Lehrerin den Klassenraum.*]$_N$ [*Fünf Minuten später verkündete sie, dass ein Test geschrieben wird.*]$_N$
- **Comment:** The presentation of the states of affairs in the text can be in the "real" order (*dann*) or in reversed order (*before*).

3.3 "Pragmatic" intra-tweet relations

This group comprises relations that serve to ease understanding for the reader (Background), try to evoke a positive sentiment (Antithesis, Concession), justify a claim of the author (Evidence, Reason), evaluate a state of affairs (Evaluation-N/Evaluation-S), and trigger actions on the side of the reader (Motivation, Enablement).

Background

- **N/S:** S increases the ability of R to comprehend an element in N. S contains orienting background information, in whose absence N would be hard to understand. In a tweet, S usually precedes N, but not always.
- **Effect:** R's ability to comprehend N increases.
- **Typical connectives:** (rarely marked by connectives)
- **Example:** [*Burkina Faso hieß bis 1984 noch Obervolta.*]$_S$ [*Nach einer EMNID-Umfrage glauben viele Europäer bis heute, dass es sich um zwei verschiedene Länder handelt.*]$_N$
- **Comment:**

Antithesis

- **N:** W has positive regard for N.
- **S:** In comparison to N, W regards the content of S as less positive or less important.

- **N/S:** N and S are in contrast (see the CONTRAST relation); because of the incompatibility that arises from the contrast, one cannot have positive regard for both of those situations. One subgroup is corrections of the kind "Not A happened, but B." In a tweet, S usually precedes N.
- **Effect:** R's positive regard for N is increased.
- **Typical connectives:** *but; while; instead; ... / aber; stattdessen; ...*
- **Examples:** *Versuch des Potsdamer Amtsgerichts, ein Grundstück zu verkaufen* [*Einmal schien es schon geglückt zu sein.*]$_S$ [*Doch der Käufer zahlte nicht.*]$_N$ (maz-6193)
  [*Ablehnung gab es nämlich nicht für den Mann,*]$_S$ [*sondern für seinen Posten, der ihn zum Befürworter des Bombodroms machen muss.*]$_N$ (maz-17300)
- **Comment:** Similar relations are CONCESSION and multinuclear CONTRAST. See Section 3.6.

Concession

- **N:** W has positive regard for N.
- **S:** In comparison to N, W regards the content of S as less positive or less important but acknowledges its truth (in a wide sense).
- **N/S:** W acknowledges that N and S in general are potentially incompatible; but in the current described situation, both hold.
- **Effect:** R's positive regard for N is increased
- **Typical connectives:** *although; but; nonetheless; still; ... / obwohl; trotzdem; aber; ...*
- **Example:** [*Sanitäre Anlagen gehören heute zum Standard großer Liegeplätze.*]$_S$ [*Doch mit der Nachrüstung tut sich Radewege schwer.*]$_N$ (maz-6488)

Evidence

- **N:** A subjective statetement or assessment that R might not believe to a degree satisfactory to W.
- **S:** A statement that R will find credible; often the description of an 'objective fact'.
- **N/S:** R's comprehending S increases R's belief of N.
- **Effect:** R's belief in N is increased.
- **Typical connectives:** (causal connectives)

- **Example:** *Streit um das Nebeneinander der Schulfächer Religion und LER* [*Und nun scheint sogar unsere Landesregierung entschlossen, diese scheinbare Gleichbehandlung der beiden Fächer zu beseitigen.*]$_N$ [*Stolpe, Reiche und Co. sagen zwar Ja zu einem möglichen Kompromissangebot aus Karlsruhe, dekretieren aber: Einen Wahlpflichtbereich LER/Religion kann es nicht geben.*]$_S$ (maz-6159)
- **Comment:**

Reason

- **N:** A subjective statement/thesis that R might not accept or regard as positive/important.
- **S:** A subjective statement/thesis.
- **N/S:** R's comprehending S increases R's readiness to accept N or to share the expressed opinion of W.
- **Effect:** R's belief in N is increased.
- **Typical connectives:** (causal connectives)
- **Example:** [*Die Militärschläge der USA verlieren mit jedem Tag weiteren Bombardements schleichend an Glaubwürdigkeit.*]$_N$ [*Mit flächenhafter Zerstörung sind weder die Taliban noch bin Laden zu treffen.*]$_S$ (maz-5701)
- **Comment:** Reason is more specific than Evidence: the difference is in whether W presents S as 'objective fact' (Evidence) or as a subjective assessment (Reason).

Reason-N

- **N:** A subjective statement/thesis.
- **S:** A subjective statement/thesis that R might not accept or regard as positive/important.
- **N/S:** R's comprehending N increases R's readiness to accept S or to share the expressed opinion of W.
- **Effect:** R's belief in S is increased.
- **Typical connectives:** (causal connectives)
- **Comment:** Dual of Reason.

Justify

- **N:** A subjective statetement or assessment that R might not accept.
- **S:** A subjective statement of a general (for example moral) attitude or position of the acting person.
- **N/S:** R's comprehending S increases R's readiness to accept W's right to present N.
- **Effect**: R's inclination for accepting W's presentation of N is increased.
- **Typical connectives:** (causal connectives)
- **Example:** [*Dieses Gericht werde ich niemals essen.*]$_S$ [*Ich bin Vegetarier.*]$_N$

Evaluation-S

- **N:** Description of a state of affairs, or a subjective statement (but not from the perspective of W).
- **S:** A subjective evaluation (positive/negative, desirable/not desirable) from W's perspective.
- **N/S:** S evaluates N.
- **Effect:** R recognizes the evaluation relationship between S and N.
- **Typical connectives:** (rarely marked by connectives)
- **Example:** [*Seine Vergangenheit schien wie ein Fluch über dem Hotelkomplex zu liegen.*]$_S$ [*Jahrelang hatte das Amtsgericht Potsdam umsonst versucht, es an den Mann zu bringen.*]$_N$ (maz-6193)
- **Comment:** In most cases the evaluating statement follows the one being evaluated.

Evaluation-N

- **N:** A subjective evaluation (positive/negative, desirable/not desirable) from W's perspective.
- **S:** Description of a state of affairs, or a subjective statement (but not from the perspective of W).
- **N/S:** N evaluates S
- **Effect:** R recognizes the evaluation relationship between S and N.
- **Typical connectives:** (rarely marked by connectives)
- **Comment:** Dual of Evaluation-S.

Motivation

- **N:** An action in which R is the actor.
- **N/S:** S provides an incentive/motivation/reason for performing the action in N.
- **Effect:** R's desire to perform action in N is increased.
- **Typical connectives:** (causal connectives)
- **Example:** [*Olympische Spiele stellen für jeden Austragungsort einen großen Gewinn dar.*]$_S$ [*Berlin muss sich jetzt um die Spiele 2016 bewerben!*]$_N$
- **Comment:**

Enablement

- **N:** An action in which R is the actor.
- **N/S:** R comprehending S increases R's potential ability to perform the action in N.
- **Effect:** R's potential ability to perform the action in N increases.
- **Typical connectives:** *so that; ... / damit; so dass; ...*
- **Example:** [*Wechseln Sie die Zündkerzen aus.*]$_N$ [*Ein Vierkantschlüssel befindet sich unter der Abdeckung.*]$_S$
- **Comment:**

3.4 Textual intra-tweet relations

The relations in this group have an "organizing" function for the tweet, they are employed by the writer to structure it. Thus, they are not in the narrow sense semantic (describing the world) nor pragmatic (targeting a change in the reader's state of mind). Notice that they apply only within individual tweets.

Attribution

- **S:** The satellite describes the fact that the author expresses the information which is presented in N.
- **N/S:** S specifies the original author of information presented in N.
- **Typical connectives:** (rarely marked by connectives; usually delimited by dash or colon)
- **Example:** Friedrich gegen doppelte Staatsangehrigkeit: [*"Wäre integrationshemmend",*]$_N$ [*sagt er #berlindirekt.*]$_S$ Mehr dazu um 19.10 http://t.co/SIqgDHHO5R

Address

- **S:** The satellite contains (a list of) syntactically independent names or @-mentions that are being addressed by the given message.
- **Effect:** R understands who is being addressed by the message.
- **Typical connectives:** *none*
- **Comment:** You should link the @-mention to the highest possible level of the discourse tree which contains information relevant for addressee.

Hashtag

- **S:** The satellite contains (a list of) syntactically independent @-mention(s). These mentions show to which user the given message is addressed.
- **Effect:** R can proceed to the information pointed by the URI.
- **Typical connectives:** *none*
- **Comment:** You should link the @-mention to the highest possible level of the discourse tree for which it still provides relevant information.

List (multinuclear)

- **N:** The nuclei provide information that play identical roles for the tweet, in the way of an enumeration.
- **Effect:** R recognizes the comparability of thenlinked items.
- **Typical connectives:** (conjunctions)
- **Example:** Was ich gestern getan habe? [*Essen kochen,*]$_N$ [*Kinder versorgen,*]$_N$ [*Bad putzen.*]$_N$
- **Comment:**

Joint (multinuclear)

- **N:** The nuclei give different kinds of information, hence there is no List enumeration; at the same time they are not in any clearly recognizable semantic or pragmatic relation. Yet there is coherence by contributing to the overall function of the tweet.
- **Effect:** R recognizes the distinct functions of the nuclei.
- **Typical connectives:** additive connectives such as *also, further / zudem, auch*
- **Comment:** Joint is to be used when a multinuclear relation is in order but none of the "stronger" relations hold.

OTHER

- **N:** Any.
- **S:** Any.
- **N/S:** The relationship.
- **Typical connectives:** *none*
- **Comment:** This relation has the lowest precedence in our relation hierarchy and should only be applied when none of the existing relations seems to be appropriate.

Preparation

- **N/S:** S precedes N in the text; S tends to make R more ready, interested or oriented for reading N. However, S does not provide information that makes understanding the content of N easier (cf. BACKGROUND).
- **Effect:** R is more ready, interested or oriented for reading N.
- **Typical connectives:** (rarely marked by connectives; sometimes by colon)
- **Example:** [*Denn eines ist klar:*]$_S$ [*Die Militärschläge der USA verlieren mit jedem Tag weiteren Bombardements schleichend an Glaubwürdigkeit.*]$_N$ (maz-5701)
- **Comment:** This relation applies when S does not have any stronger role than setting the topic of N or playing a similarly preparatory role. Thus S must not transmit a full unit of information.

Restatement

- **N/S:** S restates N, where S and N are of comparable bulk; N is more central to W's purposes than S is.
- **Effect:** R recognizes S as a restatement of N.
- **Typical connectives:** *in other words; ... / anders ausgedrückt; ...*
- **Example:** [*Der Bürgermeister hat den Stadtverordneten alle Informationen gegeben,*]$_N$ [*also gewissermaßen die Karten auf den Tisch gelegt.*]$_S$

Same (multinuclear)

- **N/S:** Used for two parts of the same segment if they were split apart by a parenthetical construction.
- **Effect:** .
- **Typical connectives:** (rarely marked by connectives; sometimes by colon)
- **Example:** [*@Schaf_reloaded*] $[da]_N$ [*wo noch Öl ist ,*]$_S$ [*riecht es noch...*]$_N$
- **Comment:**

URI

- **S:** The satellite contains a URI which is not syntactically embedded in any clause but possibly links to a resource with further relevant information.
- **Effect:** R can proceed to the information pointed by the URI.
- **Typical connectives:** *none*
- **Comment:** You should link the URI to the highest possible level of the discourse tree for which it still provides relevant information.

3.5 Inter-tweet relations

Inter-tweet relations are relations that hold between two different messages. This class subsumes all semantic and pragmatic relations that were previously defined for single tweets with the only difference that corresponding inter-tweet relations are prefixed by "r-". For example, the inter-tweet counterparts of such intra-tweet relations as Condition and Result are called r-Condition and r-Result respectively. We also allow certain textual relations (namely, Hashtag, URI, Preparation, Restatement, Joint, and List) to act as inter-tweet ties. The inter-tweet equivalents of these relations also have the prefix "r-".

Most of the inter-tweet relations defined in this scheme are mononuclear with the reference tweet usually considered as nucleus and the target tweet normally regarded as its satellite. If both reference tweet and its answer are equally important for the discussion or if they are indivisible from each other, you can also consider using one of the multinuclear inter-tweet relations. These include r-Contrast, r-Joint, r-List, r-Sequence, and r-OTHER.

In constrast to monologue texts, dialogues and multilogues often feature elements which do not serve any explicit semantic or pragmatic purposes but rather act as communicative ties. Examples of such elements include greetings, farewells, apologies, appeals, questions etc. In order to integrate such communicatve utterances into the common structure of the discourse, we have introduced a special group of so-called *communicative* relations which we present in this section.

3.5.1 Questions

For questions, we apply the same set of inter- and intra-tweet relations that we already described in Sections 3.2 and 3.3 with the only exception that the interrogative relations should be prefixed with "q-" when they are inter-tweet and with "r-q" when they denote an inter-tweet connection. For determining which relation should hold between the question and the sentence that it is addressing, we suggest the following strategy: You should temporarily replace the interrogative sentence with its (possibly contrived) answer and then look what type of discourse relation might hold between this answer and the addressed segment. Depending on which relation you choose, you simply transform it into an interrogative one by prefixing it with "q-"/"r-q" and then draw a corresponding edge between the original question and its addressed sentence. For instance, the following message/question pair can be mentally transformed as follows:

(25) – @KazuhaToyama5 Ich hab voll die Kratzer an meiner Hand :(Dank meines Haustieres -.-

(→*@KazuhaToyama5 My hand is totally scratched :(Due to my pet -.-*)

– @fr3z0r Warum hat es denn so gemacht ??

(→*@fr3z0r Why did she do that ??*)

can be transformed into

– @KazuhaToyama5 Ich hab voll die Kratzer an meiner Hand :(Dank meines Haustieres -.-

(→*@KazuhaToyama5 My hand is tatoally scratched :(Due to my pet -.-*)

– @fr3z0r Katzen wollen ja immer mit den Menschen spielen

(→*@fr3z0r Yeah, because cats always want to play with people*)

→

Here, we replaced the question "Warum hat es denn so gemacht ??" (→"*Why did she do that ??*") with its possible answer "Katzen wollen ja immer mit den Menschen spielen" ((→"*Because cats always want to play with people*"). This answer provides a cause for the situation described in the first message, namely why the pet scratched the hand of her owner. The most appropriate rhetorical relation in this case would be CAUSE, but since the original satellite is a question and also belongs to another message, we transform it into r-q-CAUSE and draw a corresponding edge between the segment $[Dank\ meines\ Haustieres]_N$ ($[Due\ to\ my\ pet]_N$) and the interrogative sentence $[Warum\ hat\ es\ denn\ so\ gemacht\ ??]_S$ ($[Why\ did\ she\ do\ that??]_S$).

3.5.2 Answers

Answers to questions should be considered as questions' satellites and linked to them via one of the following relations: R-AFFIRMATION, R-REFUTATION, R-NEI, R-INFOANSWER, or R-ALTANSWER. The particular choice of relation depends on the form and the informative content of the answer. For "yes"/"no"-answers as well as for admissions of one's inability to answer certain question, you should use R-AFFIRMATION, R-REFUTATION, and R-NEI

respectively. If an answer provides an extended information to the asked question or offers several possible options, you should use r-InfoAnswer or r-AltAnswer instead. A full description of these relations is provided below.

r-Affirmation

- **S:** S contains a positive answer or confirmation to the proposition presented in N;
- **N/S:** S confirms the veracity of the information presented in N;
- **Typical connectives:** *ja* (→ *yes*);
- **Example:**
  – [*Hast du 5 Minuten Zeit?*]$_N$
  – [*Ja.*]$_S$
  →[*Do you have 5 free minutes?*]$_N$
  –[*Yes.*]$_S$

---

r-Refutation

- **S:** S contains a negative answer, an objection to the proposition presented in N;
- **N/S:** S disconfirms the veracity of the information presented in N;
- **Typical connectives:** *nein* (→ *no*);
- **Example:**
  – [*@MyEpicReviews*] [*hast du für die Zwischenzeit ein anderes Gerät bekommen?*]$_N$
  – [*@Apfel_Bauer*] [*Nein,*]$_S$ [*aber sobald du Apple Care hast oder was dazuzahlst, bekommst du ein Ersatzgerät für die Zeit.*]
  →[*@MyEpicReviews*] [*Did you get another device for the time being?*]$_N$
  –[*@Apfel_Bauer*] [*Nope,*]$_S$ [*but as soon as you have Apple Care or are out of pocket, you can get a replacement for that time.*]

---

r-NEI (not enough information)

- **N:** a question;
- **S:** W admits his inability to answer the question presented in N;
- **Typical connectives:** *none*;

- **Example:**
  – [*Wann kommst du heute zurück?*]$_N$
  – [*Weiß noch nicht*]$_S$
  →[*When are you going to come back today?*]$_N$
  –[*Don't know yet.*]$_S$

---

r-InfoAnswer

- **N:** a question;
- **N/S:** S answers the question asked in N by providing detailed informations about the asked aspect;
- **Typical connectives:** *none*;
- **Example:**
  – [*warum?*]$_N$
  – [*@kirmiziBlume.*] [*Weil ich einen Tweet von dir gelesen habe, der mir gut gefallen hat*]$_S$
  →[*why?*]$_N$
  – [*@kirmiziBlume.*] [*Because I've read a tweet from you that I liked so much*]$_S$

---

r-AltAnswer

- **N:** a question;
- **S:** a set of possible options that all can serve as answers to the asked question;
- **Typical connectives:** *none*;
- **Example:**
  – [*Hast du was zum Schneiden?*]$_N$
  – [*@kirmiziBlume.*] [*Messer oder Schere?*]$_S$
  →[*Do you have anything for cutting?*]$_N$
  – [*@kirmiziBlume.*] [*knife or scissors?*]$_S$

3.5.3 Communicative Interaction

The rest of the inter-tweet relations serves solely communicative purposes and either controls the flow of the interaction (taking initiative) or expresses the attitude of one partner towards the statement made by her counterparts in the multilogue.

r-Apology

- **S:** W apologies for his previous action;
- **N/S:**;
- **Typical connectives:** *none*;
- **Example:**
  – [*@SabineMartiny*] [*ok, dann ist deine Ironie bei mir nicht angekommen xD*]$_N$
  – [*@heikoherberg*] [*Sorry*]$_S$ [*;)*]
  →[*@SabineMartiny*] [*ok, than I didn't get this irony xD*]$_N$
  – [*@heikoherberg*] [*Sorry*]$_S$ [*;)*]

---

r-Gratitude

- **S:** W thanks for a previous action or suggestion from R;
- **N/S:** S expresses gratitude for suggestion made in N;
- **Typical connectives:** *none*;
- **Example:**
  – [*@PatriciaDBoeni*] [*Schorle könnt ich bieten ;-)*]$_N$
  – [*@GabyRudolf*] [*Vielen Dank*]$_S$
  →edu@PatriciaDBoeni [*I could offer some spritzer ;-)*]$_N$
  – [*@GabyRudolf*] [*Thank you very much*]$_S$

---

r-Suggestion

- **S:** S makes a proposal for further action of R or further development of the discussion;
- **N/S:**;
- **Typical connectives:** *none*;
- **Example:**
  – [*Ich bin gerne ein Homosapiens!*]$_N$
  – [*@nilsding*] [*lass mal das sapiens weg :P*]$_S$
  →[*I like being Homosapiens!*]$_N$
  – [*@nilsding*] [*you should drop the sapiens :P*]$_S$

---

r-Wish

- **S:** W expresses his wishes or curses upon R;
- **Typical connectives:** *none*;
- **Example:**
  – [*#AfD: Das ist die Partei der Vollidioten*]$_N$
  – [*@kpeterlBW*] [*#Antifa verrecke nicht Deutschland*] [*verrecke du*]$_S$ [*Vollidiot*]
  →[*#AfD: It's the party of idiots*]$_N$
  – [*@kpeterlBW*] [*#Antifa doesn't kill Germany*] [*snuff it*]$_S$ [*you idiot*]

3.6 Hints on relation assignment

TODO: A TABLE THAT SUMMARIZES THE NAMES OF ALL RELATIONS, IN THEIR GROUPS

3.6.1 Connectives and relations

The following list summarizes examples of German connectives that typically signal a discourse relation. Notice that this list is by no means complete or exhaustive and all decisions still have to be made individually depending on the context.

- **aber**:
  – Antithesis;
  – Contrast;
  – Concession;
- **anschließend**:
  – Sequence;
- **anders ausgedrückt**:
  – Restatement;
- **bevor**:
  – Otherwise;
- **beispielsweise**:
  – Elaboration;
- **besonders**:
  – Elaboration;
- **da**:
  – Cause;
  – Evidence;
  – Justify;
  – Reason;
  – Reason-N;
  – Result;
- **damit**:
  – Interpretation;
  – Means;
  – Enablement;
- **dazu**:
  – Means;
- **demgegenüber**:
  – Contrast;
- **deshalb**:
  – Cause;
  – Evidence;
  – Justify;
  – Reason;
  – Reason-N;
  – Result;
- **durch**:
  – Purpose;
- **ehe**:
  – Otherwise;
- **es sei denn**:
  – Unless;
- **hingegen**:
  – Contrast;
- **mithin**:
  – Interpretation;

- **obwohl**:
  - Concession;
- **sonst**:
  - Otherwise;
- **so dass**:
  - Enablement;
- **stattdessen**:
  - Antithesis;
- **trotzdem**:
  - Concession;
- **um. . . , zu**:
  - Purpose;
- **und**:
  - Sequence;
- **weil**:
  - Cause;
  - Evidence;
  - Justify;
  - Reason;
  - Reason-N;
  - Result;

  item **wenn**:
  - Circumstance;
  - Condition;
- **während**:
  - Circumstance;
  - Contrast;
- **zuvor**:
  - Sequence;

3.6.2 The hierarchy of the RST tree

TODO: sagen dass wir per Konvention immer nur zur Wurzel des parent tweet verlinken, auch wenn vom target ein konkretes Segment angezielt wird

A central assumption of RST is that relations apply between *adjacent* text spans; there are no crossing edges in the tree. The one exception to the adjacency rule is the situation where the same nucleus is related to two or more distinct (but adjacent) satellites. TODO GIVE AN EXAMPLE, WITH PICTURE. AND WITH EXPLANATION.

In practice, these cases will be rare; instead, the satellite that is the neighbour of the nucleus will mostly give rise to a new text span, which can then in turn be related to a neigbouring span. In this way, a hierarchy originates. Since tweets are short (and often, in fact, have just one discourse segment), the RST tree will usually not be very deep. Sometimes, however, a significant degree of complexity can arise. TODO: AUTHENTIC EXAMPLE OF INTERESTING TREE.

When building hierarchy, we adhere to the "strong nuclearity principle" (Marcu, 1999): Assigning a relation to hold between two segments amounts to assigning the relation in particular to the *nuclei* of those segments. Since this rule extends to all levels of the hierarchy, it follows that the "most important" EDU of the tweet is the one that is reached by following the nucleus path from the root of the tree to its leaves. Should multinuclear relations be part of that path, more than one EDU are of equally central importance.

When establishing hierarchy through relation assignment, punctuation is usually a good indicator. But be aware that there are cases where a sentence boundary does not coincide with a boundary between larger segments. In the following example, the elliptical add-on [3] is to be added via a List relation to segment [2], and thereafter the two collectively form the nucleus of a Condition with the satellite [1].

- [*Sollte der Luftkrieg als bloße Drohkulisse gedacht gewesen sein,*]$_1$ [*verlöre er spätestens mit dem einsetzenden Winter und der unausbleiblichen humanitären Katastrophe seine Legitimation.*]$_2$ [*Und die USA ihren Rückhalt in Westeuropa.*]$_3$

In addition to these general rules for RST analysis, we establish a tweet-specific convention: Non-integrated emoticons, @-mentions, #hashtags and URIs are to be linked to the root of the tree, i.e., on highest level. If more than one of these items occur in the tweet, they are all linked as individual satellites to the root node. TODO: GIVE EXAMPLE.

3.6.3 Prefer "informative" relations

A coherence relation can be seen as the "add-on" to the sum of the meaning of the two segments. This add-on is not equally rich in information for the various relations. Consider these cases:

1. Tom bought a new car. The old one was wrecked in an accident last week.
2. Tom bought a new car. He showed it to his girlfriend.
3. Tom bought a new car. It was a Volkswagen.

In (1), the reader is likely to infer that the wrecking was the ultimate reason for Tom's purchase, hence there is a CAUSE relation between the two sentences. This is an act of interpretation that generates quite a bit of extra information. In (2), the most likely interpretation is that of temporal SEQUENCE, which is weaker and based on a very general default rule: In the absence of any indication to the contrary, two eventualities that are reported in succession will have occurred in the corresponding temporal order. Finally, in (3) we are not given two distinct eventualities; instead the second sentence merely adds an attribute to the object introduced in the first. This case of ELABORATION holds even less "extra" information, as the reader is not required to infer anything (except for resolving the anaphor, but that is an operation to occur independently of the matter of coherence relations).

TODO: REICHT DIE INTUITION FUER INFORMATIONSGEHALT, ODER MUSS DAS AUSFUEHRLICHER DARGESTELLT WERDEN?

3.6.4 Differentiating similar relations

Adversativity The first decision between the closely-related relations CONTRAST, ANTITHESIS und CONCESSION is that of mono- versus multinuclearity. When neither segment appears to be more important for the overall message, choose CONTRAST.

Among the two mononuclear candidates, CONCESSION is more specific in that it expresses the surprising correlation of two actions or states of affairs. Usually, a paraphrase test using *obwohl / although* is a reliable indicator: When this connective can join the segments with no significant change in meaning, choose CONCESSION; otherwise the appropriate relation will be ANTITHESIS, which typically links the author's subjective evaluations of the two segments.

List/Joint The central feature of LIST is the identical function of the segments for the overall message: Each nucleus of a list should contribute the same type of information for

the overall purpose. This condition does not hold for JOINT, which can be used whenever two different points are being made, which are juxtaposed with no particular relation to each other.

Since both relations are multincuclar, they of course share the general constraint that the segments must be of equal importance to the overall message – otherwise a suitable mononuclear relation is to be determined.

4 Annotation procedure

TODO: Pruefen ob man einen Entscheidungsbaum oder Flussdiagramm fr die Relationsauswahl angeben kann, als Erweiterung von Abschnitt 4.2. Wenn ja, die Deifferenzierungsinweise aus 3.6 moeglicherweise in diesen Teil integrieren.

4.1 Annotation Tool Description

After you have started `RSTTool` and loaded the annotation file in it, you should see a program window consisting of three graphical parts. The uppermost of these parts (`RSTEditor`) displays graphical nodes corresponding to the created discourse units along with the (partially) constructed RST tree for them. The middle subframe (`Reference Text Viewer`) displays the reference tweet (i.e. tweet which is being answered in the discussion), and the lowest part (`Answer Text Editor`) shows the answer to the reference tweet.

Tweets which begin the discussions and are not answers to any other tweets will be displayed in the `Answer Text Editor` first, in order to let you annotate their segments. The `Reference Text Viewer` will remain empty by that time. After you have annotated the boundaries of the EDU segments for a non-answer tweet and probably constructed an RST tree for it, you can proceed to the answers to this tweet, if any exist, by clicking on the `Next Message` button in the middle part of the program window. The non-answer tweet will then be displayed in the `Reference Text Viewer` and the answer messages will be presented in the `Answer Text Editor` consecutively one at a time.

In `Answer Text Editor`, you can create and modify discourse segment boundaries. In order to create a new boundary, you can simply click with the left mouse button right behind the word which ends the new segment. To move an existing boundary, you should press the `Control` key down and then click with the left mouse button on the boundary of the segment which you are going to edit. Keeping the key and the mouse button pressed, you can then move this boundary left and right accordingly. To delete an existing boundary completely, you should simultaneously press the combination `Control-Alt-RightMouse-Button` while the mouse cursor is pointing to the segment boundary which you want to delete.

Note that you cannot add new segment boundaries to an already annotated part of the text. You can only move and/or delete existing boundaries there. Furthermore, you cannot create or edit segment boundaries in the `Reference Text Viewer` part. In order to do so, you need to scroll the discussion back using the `Previous Message` button until the reference tweet appears in `Answer Text Editor` and then edit the boundaries there.

If these restrictions cause a significant inconvenience, feel free to contact the authors and ask them for changing this policy.

`RSTEditor` – the uppermost part of the window – allows you to construct and edit an RST tree for the created EDU segments. In order to draw an edge between two EDU nodes, you should activate the *linking mode* by clicking on the `Link` button in the top button panel (this mode is activated by default). You then can draw an edge between the two nodes by first clicking on the node which you are going to link and then clicking on the node to which you are going to connect. After that, a context-sensitive menu will appear which lets you decide if the first clicked node should be attached as a nucleus, a satellite, a part of a multinuclear relation, or as an embedded satellite or nucleus. If you are linking multiple satellites to the same nucleus, another menu will appear asking you whether the new satellite should be attached below or above the existing ones.

The next step consists in choosing an appropriate relation for the newly established edge. You will again be offered a menu with a list of possible relations for the given node constellation (i.e. the precise list of relations will depend on the nuclear-satellite relationship and whether two nodes belong to the same or to two different messages). You should click on one of the relations offered in the list to complete the edge construction. Please refer to Sections 3 and 4.2 to see how the relation type should be determined.

To change the relation type of an existing edge, you should click the `Change relation` button in the upper toolbar and then click on the child node of the respective edge. After that, the relation choice menu will be offered again and you can select a new type of the relation. To delete an edge completely, you should click the button `Disconnect` in the toolbar and then click on the child node of the relation in question.

It is highly recommended, that you regularly save your data. You can do it either by pressing the keys `Ctrl-S` (`Cmd-S` on Mac) or by selecting the menu item `File/Save` in the menubar of the window. If you see a program error while you are doing your annotation, you should better immediately stop annotating the document, save your data, and contact the authors of the guidelines in order to prevent further errors.

4.2 Segmentation and Relation Linking

Before you start annotating a single tweet, you should carefully read this message and also read its reference tweet as well, if it exists. After that, you should proceed in steps as follows:

1. Determine all propositions and proposition-like elements in the answer tweet as well as all syntactically independent @-mentions, #hashtags, and URIs and mark the boundaries of these segments in `Answer Text Editor`;

2. Temporarily strip @-mentions, #hashtags, and URIs in your mind, if these elements are not integrated in any syntactic structure within the message;

3. Determine which of the remaining propositions play the most significant informative role in the message, i.e. which proposition you would unconditionally leave if you were asked to shorten the tweet;

4. For each of the "core" propositions from Step 3, look at their adjacent left and right segments. Determine if these segments immediately relate to the core proposition or rather pertain to some other propositions back or ahead in the text. In the former case, you should link both segments together (cf. Steps 5 and 6). Otherwise, leave the core segment unmodified for the time being and repeat this step for the adjacent segment;

5. When linking two segments with a relation edge, you should proceed as follows:

   (a) Decide which of the two segments plays the more dominant role in the discourse. For doing this, you might find the ablation test useful – delete one of the two segments from the original text, replace in the remaining segment all pronouns with their respective antecedents, look if the remained text still makes sense without significantly loosing its original information. Repeat the same step for the second segment. If both segments can exist without each other, then the relation type should most probably be multinuclear. Otherwise, the dependent part should be considered as satellite, and the more independent segment as its nucleus;

   (b) Decide if the relation between two segments is semantic (i.e. describes objective events), pragmatic (i.e. describes or suggests an evaluative statement), or textual (i.e. only serves for text organization purposes);

   (c) Look at the connective between two segments or, if no connective exists, think which connective could be appropriate for them;

   (d) Find in the inverse "connective2relation" index a semantic/pragmatic/textual relation whose meaning is closest to the relationship holding between the two observed segments in text;

   (e) Check if all constraints imposed on the nucleus, satellite, nucleus/satellite relationship of this relation are satisfied in the text. Create a link, if they are. Otherwise, relax the meaning proximity of your choice and repeat the last two steps.

6. If both segment from the left and segment from the right immediately relate to the "core" segment in their middle, then you should consider if one of these adjacent segments presupposes or relies on the information from the other. Is this the case, then the segment whose information is presupposed should be linked to the middle segment first and the second one should then be linked to the new sub-tree of two segments. If both segments are conditionally independent, then they both should be linked to the core segment at the same level;

7. Look if any of the previously "mentally stripped" elements from Step 2 are adjacent to the newly merged segment and link these elements to the new segment, if they are. (multiple consecutive @-mentions, #hashtags, and URIs should previously be united to form a List);

8. Repeat Steps 4 through 8 for the newly created joined segments and their neighbours until all segments from the message were joined to a single tree.

5 Examples

TODO Provide a couple of annotation examples here, including difficult cases (e.g. from mail Wladimir December 8

6 FAQ

References

Carlson, L., Marcu, D., and Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue - Volume 16*, SIGDIAL '01, pages 1–10, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.

Marcu, D. (1999). A decision-based approach to rhetorical parsing. In Dale, R. and Church, K. W., editors, *27th Annual Meeting of the Association for Computational Linguistics, Univeristy of Maryland, College Park, Maryland, USA, 20-26 June 1999*. ACL.

Taboada, M. and Mann, W. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, 8(4):423–459.