

Who Knows What in German Drama? A Composite Annotation Scheme for Knowledge Transfer

Annotation, Evaluation, and Analysis

Melanie Andresen¹ 
Benjamin Krautter² 
Janis Pagel² 
Nils Reiter² 

1. Institute for Natural Language Processing, University of Stuttgart , Stuttgart, Germany.
2. Department of Digital Humanities, University of Cologne , Cologne, Germany.

Citation

Melanie Andresen, Benjamin Krautter, Janis Pagel, and Nils Reiter (2022). "Who Knows What in German Drama? A Composite Annotation Scheme for Knowledge Transfer. Annotation, Evaluation, and Analysis". In: *Journal of Computational Literary Studies* 1 (1). 10.48694/jcls.107

Date published 2022-12-01

Date accepted 2022-03-30

Date received 2021-12-23

Keywords

annotation, drama, knowledge, inter-annotator agreement, network analysis

License

CC BY 4.0 

Reviewers

Frank Fischer 

Fotis Jannidis 

Note

This paper has passed through the conference track of JCLS. In addition to being peer reviewed, it was presented and discussed at the 1st Annual Conference of Computational Literary Studies 2022 at the Technical University of Darmstadt

Abstract. The distribution of knowledge among characters is established as an important feature for drama analysis. Many turning points in plays are triggered by a knowledge transfer. However, knowledge transfers in plays have not yet been targeted in a formal or computational way. This paper aims at developing a framework to digitally model processes of knowledge dissemination concerning family and love relations among fictional characters in plays. We approach this as an annotation task and introduce how our composite annotation scheme models knowledge transfers among characters. We present preliminary results and discuss the question of measuring inter-annotator agreement, the calculation of which is not yet standardised for this type of annotation. Finally, we showcase an analysis of the annotated knowledge transfers on Günderröde's 1805 play, *Udohla*.

1. Introduction

"A play should lead up to and away from a central crisis, and this crisis should consist in a discovery by the leading character which has an indelible effect on his thought and emotion and completely alters his course of action," (Anderson 1965, 116) stated American playwright Maxwell Anderson (1888–1959) in an essay titled *The Essence of Tragedy* (1939). In his essay, Anderson was in search of a formula for writing a successful play. After producing a number of what he called accidentally successful plays and some box office failures, he wondered "whether or not there were general laws of governing dramatic structure which so poor a head for theory as my own might grasp and use," (Anderson 1965, 114–115) in a bid to reduce "some of the gamble [...] of play-writing." (Anderson 1965, 115) He found his answer in Aristotle's *Poetics*. To be precise, he found it in Aristotle's discussion of recognition scenes, i. e., "a change from ignorance to knowledge," (Aristotle 1995, 65) which Anderson transferred into a poetology of his own. With regard to Aristotle's remarks, Anderson characterised scenes of recognition as "essential to tragedy." (Anderson 1965, 115) He stated that a playwright must "follow the ancient Aristotelian rule: he must build his plot around a scene wherein his hero discovers some mortal frailty or stupidity in himself and faces life armed with a new

wisdom.” (Anderson 1965, 120) In Anderson’s view, then, recognition scenes, which lead to a central crisis, play a major role in shaping the course of action and the play’s impact on the audience.¹ Although we are studying recognition scenes in plays, they are a common feature not only of tragedy or drama, but of literature as a whole. They are neither limited to high, middle or low brow literature nor to certain genres or literary periods (cf. Cave 1988, 1–9). The revelation of the perpetrator in a crime novel, and how they are found guilty, can be seen as similar to recognition scenes in plays. In the following, we will limit the scope of our discussions to typically permanent relations between characters such as family relations.

An instructive example for recognition scenes, which we will use not only to illustrate the phenomenon, but also to explain our methodological approach, is Karoline von Günderrode’s two-act play *Udohla* (1805). The play revolves around effects that, according to Terence Cave, are substantial for discovery: “knowledge and the means of acquiring it, with secrets, disguises, lapses of memory, clues, signs and the like.” (Cave 1988, 2) Günderrode and her writings were virtually forgotten until Christa Wolf published selected works by her in the late 1970s (cf. Lipinski 2011, 113). *Udohla*, one of ten plays Günderrode authored, is set in a palace and its adjacent garden in Delhi. The play’s constellation of characters is presented as a “familial muddle” (Engelstein 2004, 281), i. e., family relations that are at first not transparent – neither for the audience nor for the characters appearing in the play – and later turn out to be different than expected. The play’s plot is initially marked by two important moments, both of which concern the reigning Sultan of the Mughal Empire. First, members of the Sultan’s staff, namely the vizier Mangu, the Hindu Sino, and the Dervish, argue about whether the Sultan is going to marry his recently reappeared sister Nerissa. Intrafictionally, a sibling marriage would violate Mongolian Muslim law, but not that of the hierarchically subordinate Hindu population (cf. Günderrode 1990, 204–205).² The Sultan himself is seemingly ambivalent about his desires and questions the motives of God when asking (cf. Licher 1996, 189): “Warum o Schicksal, muß ich diese lieben? / Die Einzige die du mir hast versagt.” (Günderrode 1990, 209) (“Why oh fate, must I love her? / The only one you have denied me”).³ Second, the Sultan is also told that the death sentence against Bahadar, a Hindu rebel and political traitor, has been carried out, but Bahadar’s two children have escaped. Both pieces of information have implications for the further course of the plot. Over several steps of knowledge transmission, it turns out that Nerissa is not the Sultan’s long-lost sister. Instead, she is the daughter of the previously executed Bahadar. At the same time, it becomes clear that Nerissa is the sister of the titular character Udohla. Pretending to be a relative of the Nawab,⁴ Udohla attempts to outsmart the Sultan in a bid to free his father from captivity, which – as the audience already knows – is certain to fail from the beginning. As we can see, just as in Aristotle’s prime example *Oedipus Rex*, the scenes of recognition in Günderrode’s play focus primarily on family relations.

1. Aristotle considers the recognition to be a play’s inherent counterpart to aesthetic norms of writing it. I. e., recognition as an inner-dramatic concept mirrors the demanded stringency of a tragic plot from exposition to resolution on a smaller scale (cf. Kablitz 1998, 456–457).

2. As Stefani Engelstein points out with regard to *Udohla*, “[i]t is not unusual to encounter works from the eighteenth or early nineteenth centuries which claim, falsely, that some distant culture sanctions sibling incest”. In German literature incest would oftentimes occur with a “reference to the orient and cultural hierarchies” (Engelstein 2004, 280).

3. All translations by the authors.

4. In the Mughal Empire, Nawab originally referred to an envoy of the emperor or a viceroy.

In our article, we will extend this small-scale example on the connection of family relations, the knowledge about them and a central discovery to a larger corpus of plays. For this purpose, we present a framework for the formal modelling and quantitative analysis of family-related knowledge transfers in German plays of the eighteenth and nineteenth century. By means of (manual) annotation we will operationalise⁵ knowledge transfers and thereby intertwine a content-focused approach with already established procedures of quantitative drama analysis concentrating on structural properties of theatre plays.⁶ We use annotation as a method that enriches texts or text segments with certain information, whereby the annotations takes on different functions (cf. Pagel et al. 2020, 125–141). On the one hand, we employ it to further develop and refine established quantitative methods of text analysis. This way, the annotations become part of the analysis of a play or a corpus of plays, and can support the interpretation. On the other hand, the annotations will serve as training or test data for future machine learning procedures.

In our article, we will first set forth our theoretical framework from a literary studies perspective drawing upon Aristotle's *Poetics* (section 2). Secondly, we will introduce our annotation scheme in detail (section 3). In doing so, we will illustrate how to identify text passages that include a transfer of knowledge concerning family relations and how to label them with our annotation scheme. We include knowledge changes for characters present on stage as well as the audience. Thirdly, we go on to discuss the calculation of inter-annotator agreement for our annotated data (section 4). As there is no standardised procedure yet to convincingly measure the agreement of our annotators, we will discuss some of the options and challenges when measuring agreement from a theoretical and practical point of view. Lastly, we will analyse the data we obtained during our annotation process (section 5). We will thus focus on three different perspectives. Analysing our corpus of 20 plays, we will examine at what point in drama new knowledge about family and love relations is distributed and how the internal and external communication systems are involved. Secondly, we will focus on one key piece of information and present a visualisation of knowledge flow based on our annotations. Our third perspective concentrates on a methodological question: Can we use our annotation data to employ new, more content-based ways of literary network analysis? Can this approach help in identifying important characters for the play's action and contribute to improving the integration of quantitative network analysis with qualitative close readings, thus bridging the perceived gap of quantitative and qualitative methods (cf. Mueller 2012)? We will discuss these questions with regard to Günderrode's *Udohla*.

2. The Distribution of Knowledge in Plays

The interplay of internal and external communication systems in drama, i. e. the communication of the fictional characters on the one hand and the perception of this com-

5. For our understanding of operationalisation cf. Pichler and Reiter 2021, 1–29 and Pichler and Reiter 2020, 46–47

6. Research focusing on these structural properties includes formally analysing character speech (cf. Reiter and Willand 2019 or Krautter and Willand 2021, 111–118), examining the distribution of characters within a play or a corpus of plays (cf. Marcus 1973 [1970]; Yarkho 2019 [1935–1938]) and network analysis (cf. Moretti 2011; Trilcke 2013; Trilcke et al. 2015).

munication by the audience on the other, is considered one of the central “qualities necessary for identifying dramatic communication” (Pfister 1988, p. 49). As Bernhard Asmuth points out in his introduction to drama analysis, a play as a whole is not only a sequence of actions, but also a multi-perspectival processing of knowledge (cf. Asmuth 2016, 114).⁷ In the light of events that may have taken place before the actual plot of the play’s main text, characters are – potentially – already set apart from each other by a different degree of knowledge. Herein, we employ a broad understanding of knowledge that is not strictly limited to the classical notion of propositional knowledge as “justified true belief” (e. g., Pollock and Cruz 1999, 13 or Ichikawa and Steup 2018) which originated from Plato.⁸ As it is not uncommon for literature to deliberately play with knowledge, facts, beliefs, hearsay, and rumours,⁹ we opt for a more “lightweight sense of knowledge” (Ichikawa and Steup 2018). In our case, this includes beliefs that are both justified and depicted to be true, but might later turn out to be false, e. g., through scenes of recognition.

A character’s level of knowledge can change continuously in the course of the play. At the same time, the relationship between the audience’s level of information and that of the individual characters in the play is constantly adjusted. The exposition, e. g., reduces the knowledge gap between the audience and the characters that prevails at the beginning of a play (cf. Asmuth 2016, 122). The disparities in the “levels of awareness” (Pfister 1988, 49) can be attributed primarily to two causal differences between the internal and external communication systems: While the audience in its observer role perceives every scene of the play and can thus compare and aggregate partial knowledge of the characters, it sometimes remains unclear what prior knowledge the characters actually have. This also applies to possible time leaps, for instance between two acts of the play. Furthermore, it might not be clear to what extent the statements of a character correspond to the “facts” of the fictional world, i. e., whether the statements are credible (cf. Jeßing 2015, 50–51). Depending on the course of the plot, the audience can either have an information advantage or an information disadvantage over the characters acting on stage at different times. The relative level of being informed between the audience and a character can change from scene to scene. The same applies to the internal communication system of the plays’ characters, when comparing the degree of knowledge different characters have in a certain scene. For this phenomenon, Bertrand Evans coined the term “discrepant awareness” (Evans 1960, p. VIII). This “discrepant awareness” between two characters can thus lead to rather different evaluations of the same action or situation. If we think of Günderrode’s *Udohla*, a character’s judgement of the supposed marriage between the Sultan and Nerrisa would greatly depend on whether the character knows that the Sultan and Nerrisa are not siblings and on the character’s religious views, i. e., them being Hindu or Muslim. In this situation, the lack of knowledge or a perceived, but actually incomplete awareness will influence the judgement in one way or the other.

The gap between the characters’ level of knowledge and that of the audience can be seen as an important element of suspense in drama, as it ensures sustained attention and

7. As we have already illustrated by the example of *Udohla* and its portrayal of Hindu culture not condemning sibling incest, intra-fictional knowledge does not have to be valid outside the represented fictional world.

8. Defining knowledge as ‘justified true belief’ is controversial in itself (cf. Gettier 1963 and Dutant 2015).

9. The plays of Heinrich von Kleist are a prominent example of failed communication between characters creating rumours that are believed to be true (cf. Dubbels 2012).

emotional excitement (cf. Anz 2007, 464). This applies to both the suspense felt when one is curious about *what* is going to come up next and the suspense arising in respect to *how* something that is already known to be happening is going to happen.¹⁰ In this respect, the device of dramatic irony is particularly important, as it is based precisely on this gap of being informed. The audience's knowledge advantage with respect to an upcoming action is, thus, a prerequisite for dramatic irony. In understanding a remark that is innocuous from the perspective of the speaking character, the audience can interpret the utterance as an allusion to the catastrophe that is later actually realised.¹¹ Consequently, elements such as dramatic irony are closely linked to the play's effect on the audience: Is the play supposed to convey a moral theorem? Is it meant to purify the audience's affects? Should it educate the audience? Or is it simply meant to entertain? In his *Poetics*, Aristotle defines the (cathartic) effect as the central concern of tragedy.¹² He considers reversal (peripeteia) and recognition (anagnorisis) as important building blocks to evoke pity and fear, the desired affects caused by a tragedy.¹³ Recognition is directly related to Evans' concept of "discrepant awareness," for Aristotle defines recognition as "a change from ignorance to knowledge, leading to friendship or to enmity, and involving matters which bear on prosperity or adversity." (Aristotle 1995, 65) Since such scenes of recognition are ideally linked to the reversal, i. e., "a change to the opposite direction of events" (Aristotle 1995, 65), they represent central moments of knowledge transmissions that can be decisive for understanding and interpreting a play. The examples Aristotle used to illustrate recognition and reversal "are taken solely from the field of familial *philia*" (Destrée 2020, 117). This is one reason why our annotation experiments focus on knowledge about family relations.

3. Annotating Knowledge Transfers

The aim of our research is to model knowledge transfers in German plays by means of annotation. This enables us to empirically analyse the textual implementation of the theoretical considerations described in the previous chapter. While knowledge is a broad phenomenon, we restrict our annotation to the domain of knowledge about familial character relations. We employ a wide understanding of knowledge that does not imply that the information at hand must be correct. We therefore also include beliefs. In this section, we will present our current annotation scheme. We developed the guideline by annotating 16 plays in the course of roughly a year. The full (German) guideline as used by our annotators is available online.¹⁴ The annotation is performed using the tool CorefAnnotator¹⁵ (Reiter 2018).

Our annotation scheme targets text sections in which knowledge transfers take place.

10. See DiYanni (2000, 22): "One of our main sources of pleasure in plot is surprise, whether we are shown something we didn't expect or whether we see *how* something will happen even when we may know *what* will happen. Frequently surprise follows suspense – fulfilling our need to find out what will happen as we wait for a resolution of a play's action."

11. Contrary to what this wording suggests, dramatic irony is not limited to tragedies, but is often found in comedies as well.

12. There are numerous studies that examine Aristotle's mention of catharsis in great detail. Cf. for instance Schmitt (2008, p. 333–348 and 476–510).

13. There is a great debate about what the two affects mentioned by Aristotle actually express and how to translate them properly (cf. Schadewaldt 1955, 129–171).

14. See: <https://doi.org/10.5281/zenodo.5729706>.

15. See <https://doi.org/10.5281/zenodo.1228105> for stable release versions and <https://github.com/nilsreiter/CorefAnnotator/> for development versions.

More precisely, we annotate a text section if:

- a) the knowledge concerning character relations of at least one of the characters or the audience is changed, or
- b) a character's knowledge about the knowledge of another character is changed.

A case of a) would be a text section in which character A learns that B and C are siblings. An example for b) is a section in which B learns that A knows that B and C are siblings. The latter can be understood as knowledge about knowledge, or meta-knowledge.

Annotation spans are not fixed to a specific length. Knowledge transfers can happen in one sentence or even a word, but can also be extended over a whole paragraph, especially when knowledge is distributed implicitly. However, our annotators are encouraged to identify a span that is as short as possible. When a relevant text span is identified, it is annotated with a label that uses this pattern:¹⁶

(1) transfer(SOURCE, TARGET, KNOWLEDGE, ATTRIBUTES)

The SOURCE is usually a character that provides a piece of information, but can also be an object or an action that allows for inferences about character relations, for instance when Saladin recognises the handwriting of his brother in Lessing's *Nathan der Weise* (1779). The TARGET is always a character or a group of characters (and/or possibly the audience) whose knowledge is changed. The item KNOWLEDGE is restricted to knowledge about character relations and, more precisely, to the set of relations presented in Table 1. Optionally, ATTRIBUTES can be added, for instance to mark the information as a lie or as uncertain. The latter is especially frequent as many dramatic texts play with strong allusions to a fact that is ultimately confirmed only at the end.

In *Udohla*, by the playwright Karoline von Günderrode, the vizier Mangu lets the audience know that Nerissa is the Sultan's sister (which turns out to be wrong). This is annotated as follows:

(2) transfer(mangu, audience, siblings(sultan, nerissa))

Characters are referenced by the identifier they receive in the *Drama Corpora Project* (DraCor, Fischer et al. 2019). Characters that do not have dialogue, do not have such an identifier. Instead, they are given an identifier by our annotators. Frequently, characters are not introduced by name and their identity is (partly) unclear. We annotate such character mentions as a variable in capital letters. In the play *Magie und Schicksal* (1805) ("Magic and Destiny") by Günderrode, the character Cassandra mentions a son whom the audience did not hear about before. At first, we do not have any additional knowledge about this son and therefore annotate him as a variable:

(3) transfer(cassandra, audience, parent_of(cassandra, CHILD[CASSANDRA]))

Later in the play, it is revealed that the character Ligares is in fact the mentioned child of Cassandra. We can now annotate that the variable CHILD[CASSANDRA] and Ligares are identical:

16. This notation is inspired by the syntax of the programming language Prolog.

(4) `transfer(cassandra, audience, identity(CHILD[CASSANDRA], ligares))`

Note that it is also possible to fill any of the positions in the annotation label with a list of several characters by enclosing them in square brackets. This is used extensively, for instance, when Nerissa (in *Udohla*) reveals in the final scene that she is the daughter of the Sultan's enemy Bahadar (who was just killed on his order):

(5) `transfer(nerissa, [sultan, udohla, mangu, sino, audience],
child_of(nerissa, "Bahadar"))`

As mentioned above, we restrict our annotations to the domain of knowledge about character relations, i. e. family and love relations. Table 1 gives an overview of all character relations that are included in the annotation scheme. Formally, we differentiate directed relations such as `parent_of(PARENT, CHILD)`, where the position of characters is important because of the asymmetry of the relation, from undirected relations. When annotating undirected, symmetric relations such as `siblings(SIBLING-A, SIBLING-B)`, the order of the characters is irrelevant. Semantically, the relations form three groups, the biggest of which are family relations and love relations.¹⁷ The last group of identity relations is not about relations between two characters in the strict sense, but includes a) cases where we learn a (first, or additional) name of a character, and b) cases where two characters are revealed to be the same, as in example 4. All relations in the table can be negated by adding a ! at the beginning, e. g., `!siblings(nerissa, sultan)` to express that Nerissa and the Sultan are not siblings.

	Directed Relations	Undirected Relations
Family Relations	<code>parent_of(PARENT, CHILD)</code> <code>child_of(CHILD, PARENT)</code> <code>aunt:uncle_of(AUNT:UNCLE, NIECE:NEPHEW)</code> <code>niece:nephew_of(NIECE:NEPHEW, AUNT:UNCLE)</code>	<code>siblings(SIBLING-A, SIBLING-B)</code> <code>cousins(COUSIN-A, COUSIN-B)</code> <code>relatives(RELATIVE-A, RELATIVE-B)</code>
Love Relations	<code>in_love_with(LOVER, TARGET)</code> <code>widow:er_of(WIDOW:ER, DEAD-PARTNER)</code>	<code>lovers(LOVER-A, LOVER-B)</code> <code>couple(PARTNER-A, PARTNER-B)</code> <code>engaged(PARTNER-A, PARTNER-B)</code> <code>spouses(PARTNER-A, PARTNER-B)</code>
Identities	<code>has_name(A, NAME)</code>	<code>identity(A, B)</code>

Table 1: Character relations covered by our annotation scheme. Where applicable, the prefixes `grand-`, `step-`, `foster-`, `god-` and `ex-` as well as the suffix `-in-law` can be added.

While the annotation guideline covers most of the knowledge transfers happening in the plays, some challenges remain. This is related to the rather simplistic communication model that underlies the scheme. Although we try to include rules of pragmatic communication in our annotation decision, we formally conceptualise knowledge distribution as transfer: The knowledge of one character is transferred to another character. This is motivated by the structure of plays, which is characterised by the alternation of character speech. This view assumes, however, that the communicated information is understood in the intended way. While this might be true in many cases, there are exceptions. There can be misunderstandings, pieces of information can be interpreted in different ways, and different prior knowledge or values can influence the understanding. Currently

¹⁷ The group of love relations is very heterogeneous and subsumes all relationships motivated by love, sexual or material interest, which one might want to differentiate in follow-up studies.

No	Author	Text
1	Johann Wolfgang Goethe	<i>Iphigenie auf Tauris</i>
2	Johann Wolfgang Goethe	<i>Die natürliche Tochter</i>
3	Johann Wolfgang Goethe	<i>Stella</i>
4	Franz Grillparzer	<i>Die Ahnfrau</i>
5	Friedrich Hebbel	<i>Maria Magdalene</i>
6	Hugo von Hofmannsthal	<i>Elektra</i>
7	Hugo von Hofmannsthal	<i>Der Rosenkavalier</i>
8	Heinrich von Kleist	<i>Die Familie Schroffenstein</i>
9	Friedrich Maximilian Klingler	<i>Die Zwillinge</i>
10	Jakob Michael Reinhold Lenz	<i>Der Hofmeister</i>
11	Gotthold Ephraim Lessing	<i>Nathan der Weise</i>
12	Gotthold Ephraim Lessing	<i>Emilia Galotti</i>
13	Johann Gottlob Benjamin Pfeil	<i>Lucie Woodvil</i>
14	Friedrich Schiller	<i>Die Braut von Messina</i>
15	Friedrich Schiller	<i>Die Räuber</i>
16	Arthur Schnitzler	<i>Komtesse Mizzi</i>
17	Luise Adelgunde Victorie Gottsched	<i>Das Testament</i>
18	Karoline von Günderrode	<i>Udohla</i>
19	Karoline von Günderrode	<i>Magie und Schicksal</i>
20	Johanna von Weißenthurn	<i>Das Manuscript</i>

Table 2: Name and author of all plays that are part of the annotated corpus. The first 16 have been annotated in the process of guideline development, the last four have been used for agreement calculation. The latter group is being continuously expanded.

we also assume that the communicating characters are transparent to all characters involved. This is not true for text passages where characters transfer knowledge to a character whose identity is unclear to the speaker. For instance, in Schiller's *Braut von Messina* (1803), Don Cesar confesses his love to Beatrice without even knowing her name – not to mention that she is also his brother's lover and his sister. The annotation shown in example 6 captures the view of the audience but does not conform to the perspective of Don Cesar. While the guidelines do provide solutions for such scenes, we want to improve their generalisability in future versions, once more texts with similar constellations have been annotated.

```
(6) transfer(don_cesar, [audience, beatrice], in_love_with(don_cesar,
    beatrice))
```

All plays are annotated by two student annotators independently, then all deviations between the two versions are discussed with one of the authors. Afterwards, each of the annotators produces a revised version. Contrary to many other annotation projects, the aim of this step is not to create one consensus version of the annotation. The annotation task is complex and many text passages can be interpreted in more than one way. In addition, the annotation scheme sometimes allows for different ways of modelling a knowledge transfer. Therefore, the revision focuses on plausibility, consistency and formal correctness of the annotations. However, additional consensus versions were created for analyses that require one single reference version, because the focus does not lie on annotation variation (as in section 5.3).

Table 2 gives an overview of the current state of our corpus of annotated plays. Our choice of texts was guided by the goal to cover a broad range of phenomena, as this will

allow us to:

1. create an annotation guideline that covers as many textual phenomena as possible and is largely applicable to unseen texts without requiring additions and modifications, and
2. in the future, work on the automation of the annotation and train a robust machine learning model that is able to generalise across different epochs, authors, and genres.

We therefore combined plays in which we knew that knowledge about character relations is important for the plot (based on prior readings and secondary literature) with canonical plays with a less obvious focus on family relations. We included tragedies as well as comedies and did not restrict ourselves to a specific epoch, but annotated a broad mix of plays from the eighteenth and nineteenth century.

In the process of developing the annotation guideline, we annotated 16 dramatic texts. Once the guidelines were consolidated, we began tracking the initial versions of our annotators for the calculation of inter-annotator agreement. The annotation of this second round is ongoing. We have chosen four of these plays to develop a suitable way of determining inter-annotator agreement for a complex annotation task as this. The next section will present and discuss our current measure.

4. Calculating Inter-Annotator Agreement

For manual annotation and coding tasks in a wide range of disciplines, measuring inter-annotator agreement (IAA, sometimes also called “inter-coder reliability”) is a standard procedure (cf. Artstein and Poesio 2008; Krippendorff 2004), much like the evaluation of automatic predictions based on machine learning. The goal of this metric is to have a quantitative view on the agreement between annotators, and ultimately to evaluate the quality of the annotation guidelines, the annotation process, or the annotations themselves. Unlike an evaluation of automatic predictions, there is no ‘gold standard,’ i. e., no set of annotations is considered to be true. Instead, IAA ‘only’ measures the agreement. A corner stone of IAA metrics is to take into account expected agreement (also called ‘chance agreement’), i. e., agreement that is achieved by random annotation decisions. This is done to compensate for the difficulty of the task; if there are more classification categories, the task is considered to be more difficult because there are more options to choose from, and the expected agreement decreases. On structurally simple tasks such as part-of-speech tagging, measuring IAA is well established and understood: Fleiss’ Kappa (Fleiss 1971), for instance, can be used to calculate the IAA between n annotators, who assigned one of k categories to each of N items.¹⁸

The annotation task we discuss in this article, however, is more complex: i) Annotation decisions are not made in isolation, but depend on the textual context as much as on previously made decisions. As we are only annotating the transfer of new information to the target, a subsequent mention of the same information by and to the same character

18. For an overview of annotation metrics that is tailored to readers in computational literary studies, see also Reiter and Konle (2022).

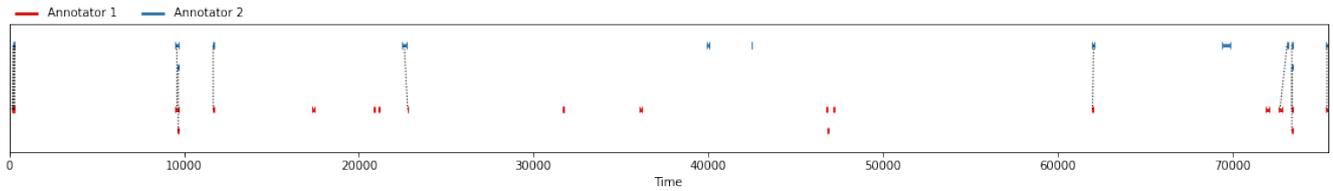


Figure 1: An alignment between the annotations in Günderrode's *Magie und Schicksal* as established by Gamma.

is not annotated. Consequently, each annotation label may only appear once in a text.¹⁹ ii) After having decided that a knowledge transfer takes place (and selecting the exact boundaries), annotators need to make decisions about the SOURCE and TARGET of the transfer, the participants of the character relation and its direction, and, finally, about potential attributes of the annotation (e. g., the transfer being a lie). iii) The annotation is not done on fixed, pre-defined units, but the annotation spans can be defined freely. All three properties make measuring IAA difficult.

The metric Gamma (Mathet et al. 2015) has been proposed as a versatile, highly adaptable metric for various tasks. It has several properties that make it promising for our use case: i) To calculate expected agreement, it samples a large number of random annotations from the existing annotations. Based on these random annotations, we can compute expected agreement in the same way we calculate observed agreement. This way, expected agreement can be measured empirically instead of theoretically, which makes it less dependent on assumptions and more widely applicable. ii) Equality between annotation categories can be graded. Instead of only recognising that $\text{transfer}(X, Y, \text{parent_of}(P, C))$ is different from $\text{transfer}(X, Z, \text{parent_of}(P, C))$, we can provide a function to express the similarity of the two annotations as a value between zero and one. This allows us to define the similarity of the annotations above to be less than one, but larger than zero. iii) For measuring observed agreement, Gamma first establishes an alignment between the different annotators' annotations. This alignment can also be visualised and inspected, which is a helpful tool in the annotation process. Figure 1 shows an example for the established alignments. The overall Gamma score is calculated based on pairwise similarity functions between two (or more) annotations that are aligned. Since Gamma is computed over disagreements instead of agreements, we will discuss the calculation of disagreements in the following.²⁰ The final Gamma score, however, can be interpreted in the same way as other metrics: The higher the score, the better the agreement.

The final Gamma value is a weighted combination of two aspects of disagreement. Positional disagreement expresses how different the annotations' positions are, while categorical disagreement compares the labels that the annotators have assigned. The exact calculation of positional and categorical disagreement, as well as the weighting of these two components, can be customised. The two values are not fully independent of each other, as the alignment of the annotations takes the labels into account, i. e., Gamma attempts to align annotations with the same label.

19. A possible exception are love relations, where characters might change their mind several times and thus confess their love to someone more than once.

20. While this distinction is mathematically important, conceptually it is not, because we can always convert an agreement score into a disagreement score by subtracting it from 1.

4.1 Gamma Setup

To calculate Gamma, we use the *pygamma-agreement* implementation²¹ with the CBC solver. To adapt Gamma to our purposes, we have defined custom functions for categorical and positional disagreement.

For the positional dissimilarity, we consider each annotation that overlaps by at least one character²² as having the same position. Annotations that do not overlap become more dissimilar with increasing distance.

For the calculation of categorical disagreement, which is defined for a tuple of annotations (u and v , one from each of the two annotators), we look at the six components of the annotated predicate separately: Those are SOURCE, TARGET and ATTRIBUTE of the knowledge transfer as well as the KNOWLEDGE, composed of literary characters 1 and 2 involved in the relation and the relation name. The disagreement d for each of these components is combined linearly, allowing us to focus on each of them individually by giving them a weight w (Equation 1).

$$\begin{aligned}
 d_{\text{cat}}(u, v) &= w_{\text{source}}d_{\text{source}}(u, v) & (1) \\
 &+ w_{\text{target}}d_{\text{target}}(u, v) \\
 &+ w_{\text{attribute}}d_{\text{attribute}}(u, v) \\
 &+ w_{\text{character 1}}d_{\text{character 1}}(u, v) \\
 &+ w_{\text{character 2}}d_{\text{character 2}}(u, v) \\
 &+ w_{\text{relation name}}d_{\text{relation name}}(u, v)
 \end{aligned}$$

The dissimilarity of the individual components is calculated in different ways. The components' relation name and attribute are always single values that can be directly compared, returning a value of 0 or 1. For the components containing characters (i. e., SOURCE, TARGET, character 1 and character 2), annotators can express lists of characters, and they make use of this frequently (see Example 5). For this reason, we use the Jaccard distance (Jaccard 1912) as a measure of dissimilarity between the two lists (Equation 2). This distance is calculated as the inverse of the Jaccard similarity, which measures how many of the elements that appear in at least one of the lists (their union) are present in both list (their intersection), resulting in a value of 1 if the two lists are identical.

$$d_{\text{target}}(u, v) = 1 - \frac{|u_{\text{target}} \cap v_{\text{target}}|}{|u_{\text{target}} \cup v_{\text{target}}|} \quad (2)$$

The Jaccard distance is also employed to measure dissimilarity between character groups for undirected relations. If both annotations specify an undirected relation, we compare the entirety of characters by Annotator 1 with the entirety of characters by Annotator 2.

Once the categorical and positional dissimilarity are calculated, they are weighted

21. See: <https://github.com/bootphon/pygamma-agreement>.

22. In this case, 'character' refers to the graphic symbols of a text, not the literary characters.

Text	Components of the annotated predicates							
	Position	Relation						All
		only	Source	Target	Attribute	Char. 1	Char. 2	
Gottsched: <i>Das Testament</i>	0.403	0.331	0.414	0.400	0.295	0.326	0.243	0.250
Günderrode: <i>Magie und Schicksal</i>	0.525	0.582	0.526	0.521	0.417	0.369	0.507	0.392
Günderrode: <i>Udohla</i>	0.454	0.356	0.246	0.416	0.144	0.199	0.241	0.146
Weißenthurn: <i>Das Manuscript</i>	0.623	0.606	0.476	0.599	0.510	0.488	0.518	0.508

Table 3: IAA scores for Gamma, when various components are taken into account. In column **Position only**, categorical agreement is irrelevant. Column **All** shows scores when all components are uniformly weighted ($\frac{1}{6}$).

against each other in order to receive the final Gamma score based on the total dissimilarity shown in Equation 3. Since we are generally more interested in the categories, we set $\alpha = 1$ and $\beta = 2$, thus categorical disagreement is twice as important as positional disagreement. In addition, we take into account the fact that we measure positional disagreement over (typographic) characters instead of tokens (as in the original version of Gamma) and that utterances may be a more relevant unit than tokens. Thus, we set α_0 to 0.001. Note that weighting is a decision without a neutral option and any choice will be debatable.

$$d(u, v) = \alpha d_{\text{pos}} + \beta d_{\text{cat}} \quad (3)$$

4.2 Inter-Annotator Agreement Results

Table 3 shows Gamma scores for four texts, using different ways of weighting positional and categorical disagreements and of comparing the predicates used in the annotation. For the first column, ‘Position only’, we set the weight of the categorical agreement to 0, such that the score only depends on the positional agreement and two annotations are considered similar if they occupy the same position, irrespective of their categories. The next six columns evaluate one component at a time, with a weighting of 0.95 of the component of interest, and 0.01 for the other five components.²³ The final column, ‘All’, shows a score for which all components are considered with a uniform weight of $\frac{1}{6} = 0.166$. As discussed above, the scores are calculated on the best possible alignment, which is determined by the Gamma metric itself. This means that every column in Table 3 is (potentially) calculated with a different alignment.

If these scores are evaluated in usual IAA terms,²⁴ they are rather low. Even relatively clear components, such as the source of the transfer (which is often just the character speaking), seem to be more difficult than expected. The variance between texts is also noticeable. Günderrode’s *Udohla* seems to be the most difficult one to annotate, while the results for Weißenthurn’s *Das Manuscript* are much more promising.

23. The decision not to set the weights to 0 and 1 was made after inspecting some of the alignments that Gamma produced. By specifying a small weight for each component, each component has some influence on the established alignments, and we prevent an alignment that is only based on a single component.

24. Many publications at this point refer to the table by Landis and Koch, published in the context of diagnostics of multiple sclerosis diagnosis, but even Landis and Koch consider the table “arbitrary” (Landis and Koch 1977, 165).

The main reason for the low scores, however, is not a disagreement on individual components of the knowledge transfer, but the fact that many annotations do not have a counterpart at – roughly – the same position in the text (as can also be seen exemplary in [Figure 1](#)). This means that many of the annotations are aligned with a dummy annotation which yields maximal categorical dissimilarity. Thus, it seems to be more difficult to decide where an annotation should be made than to decide on the individual annotation's categories.

4.3 Discussion

The calculation of inter-annotator agreement for complex annotation tasks like the one we have presented here is not straightforward. To tackle this issue we decided to use the highly adaptable measure Gamma. Our customised version of Gamma allows for a tentative assessment of the agreement between the two annotations. It permits us to evaluate the difficulty of annotating a play compared to other plays. In addition, we get a clearer picture regarding the difficulty of the annotations' different components (like SOURCE vs. TARGET). However, many properties of the annotations are not yet captured in a fully satisfactory way and the highly adaptable nature of Gamma presents us with a large number of choices, not all of which can be motivated theoretically.

The core conceptual question is what to consider as agreement (or disagreement). Two annotators marking the exact same span of text with the exact same label is not very likely and not necessary. We decided to consider two annotations as an agreement if the annotation spans overlap, because there is usually some key term (like *mother*) that will definitely be annotated while the question of how much syntactic context should be included will be answered differently by different annotators. For some cases, we could go even further and declare two annotations an agreement if they appear in the same scene or act. For love relations that develop gradually, finding agreement on which text segment is crucial for knowing that A loves B is especially difficult. One annotator might interpret the first allusions as justified evidence for an annotation (see example 7 from Weißenturn's *Das Manuskript*) while another might wait for a segment that removes the final doubt (example 8). Both decisions can be legitimate, and a contrasting analysis of how different readers perceive the development of the relationship could be very fruitful. Our current agreement measure does not account for this scenario, however.

- (7) EMERIKE etwas verschämt. (*a little shy*)
 Ich kenne einen Andern, den ich gerne glücklich machen möchte.
I know someone else, whom I would like to make happy.
 FLINT.
 Einen – Andern?
Someone – else?
 EMERIKE. Ja – ich kenne – [...], denn ich möchte Ihnen sagen – Herzlich. daß ich Ihnen recht gut bin.
Yes – I know – [...], because I want to tell you – Sincerely. that I am quite sympathetic to you.
- (8) EMERIKE mit einem Blick auf Flint. (*looking at Flint*)
 Ach nein! er will mich nicht, und ich werde doch keinen Andern lieben.
Oh no! He does not want me and I will still not love anyone else.

With regard to the comparison of annotation labels, we also want to incorporate inferences that can be drawn from relations that are logically related or equivalent. For undirected relations, it is obvious, e. g., that `siblings(A, B)` and `siblings(B, A)` are semantically equivalent. As described above, this is taken into account by our customisation of Gamma. But more complex cases would need to be covered as well. Directed relations oftentimes have a complimentary relation that can be used to express the same fact, such as `parent_of(A, B)` and `child_of(B, A)`. Our annotators are asked to base their decision on the textual expression of the relation, but some ambiguities remain. Depending on previous knowledge about familial character relations, other pairs of relations can also be equivalent. In *Die Familie Schroffenstein* by Heinrich von Kleist, we encounter such an ambiguity in the list of characters at the beginning of the play:

- (9) Rupert, Graf von Schroffenstein, aus dem Hause Rossitz.
Rupert, count of Schroffenstein, from the house of Rossitz.
 Eustache, seine Gemahlin.
Eustache, his wife.
 Ottokar, ihr Sohn.
Ottokar their/her son.

The pronoun *ihr* can either be plural or singular, feminine, and thus refer to Eustache and Rupert or only to Eustache. This corresponds to the following two options for the annotation:

- (10) `transfer("Dramatis Personae", audience, child_of(ottokar, eustache))`
 (11) `transfer("Dramatis Personae", audience, child_of(ottokar, [eustache, rupert]))`

Given that we know that Rupert and Eustache are married, we might want to consider these annotations a match, even though the surface form is different. To actually compare the readings of the two annotators, we would need to analyse if one reading is semantically equivalent to the other. We are therefore working on an inference system that automatically expands the annotated relations to all relations that are logically inferable. Once this is completed, we can update our notion of agreement and consider annotations as agreeing if they result in the same knowledge base for the characters involved. This is complicated by the fact that, in example 9, strictly speaking, we cannot logically infer that Rupert is Ottokar's father. Still, a human reader of this list will most likely assume this relationship unless presented with contradictory information.

For the implementation of Gamma, the choices of weighting need to be further discussed and refined. Fundamentally, it is necessary to justify how positional agreement and categorical agreement should be weighted against one another. As our annotation labels are complex, we additionally must establish a weighting of the individual components. In our current implementation, all six components are considered independently and have equal weight. This independence assumption raises new questions, however. Currently, the way we compare related characters is determined by the relation name. If both annotators employ `siblings` as a label, we can compare the characters with the Jaccard index. It is unclear, however, how to proceed if one annotator specifies a directed and the other an undirected relation. In addition, the components' independence can lead to non-intuitive judgements. If one annotator argues for a given text passage that

	Annotation 1	Annotation 2
No. of transfers	551	506
Average no. of transfers per play	26.24±11.58	26.63±11.65
No. of transfers per 1000 tokens	1.11±0.54	1.01±0.58

Table 4: The number and density of transfers for both annotations.

parent_of(A, B), whereas the other annotator argues that loves(A, B), this would be considered a 2/3 match, even though the transmitted information differs significantly.

5. Analysing Annotated Knowledge Transfers

The following section is dedicated to analysing our annotated corpus, with a focus on three different aspects of our annotations, as a showcase of how our annotations can be further used. As a first investigation, we concentrate on the quantitative properties of the annotated relations and provide an overview of the annotations (5.1): How many knowledge transfers are annotated per play? Which character relations are the most frequent? And when in a play is knowledge distributed in the inner and outer communication system? We discuss the results with regard to established views in drama theory. Second, we focus on how one key piece of information is distributed among the characters, and present a new visualisation for such knowledge flows (5.2). Third, we further explore the potential of the annotations in a network analysis of G nderrode’s *Udohla*. Dramatic network analysis is currently based mostly on so-called configurations (cf. Pfister 1988, 171–176). Using a more content-based form of character networks by exploiting our annotations, we attempt to chart a path to better integrate quantitative analysis and interpretative reading. We thus not only visualise the annotated knowledge transfers as a network, but also compare different characters in view of centrality measures (5.3). As we have argued in the previous section, there can be more than one way of interpreting a text and possibly also more than one way of modelling knowledge transfers in our annotation scheme. While we use both versions of the annotations for our statistical analyses in Section 5.1, we have created a consensus version for the analysis of G nderrode’s *Udohla* for simplicity.

5.1 Quantitative Overview of the Annotations

In total, our analysed corpus consists of 20 plays (see Table 2 for an overview) annotated by two annotators. It contains 551 transfer annotations for Annotator 1 and 506 transfer annotations for Annotator 2. Averaged over both annotations, there are 26.4 (± 11.5) annotations per play and 1.06 (± 0.56) annotations per 1000 tokens. The standard deviations indicate a substantial variation between the plays. Table 4 shows how these values are distributed in detail. Table 5 displays the ten most frequently annotated relations for each annotator. Overall, the ranking is fairly similar for both annotators, and two adjacent relations switch ranks only twice. The relation *in_love_with* is by far the most frequent, with its negation following shortly after. In contrast to most family relations, love relations can change over time. They can be hinted at, be the content of rumours, or trigger an important conflict for a play’s rising action. Hence, they are talked about more often than other relations. The identity relation occupies the

Annotation 1		Annotation 2	
Relation	Count	Relation	Count
in_love_with	114	in_love_with	112
child_of	72	identity	74
identity	72	child_of	54
parent_of	43	parent_of	45
!in_love_with	42	has_name	44
has_name	36	!in_love_with	36
engaged	32	engaged	32
siblings	26	siblings	25
lovers	16	spouses	16
spouses	16	lovers	11

Table 5: Ten most frequently annotated relations per annotator.

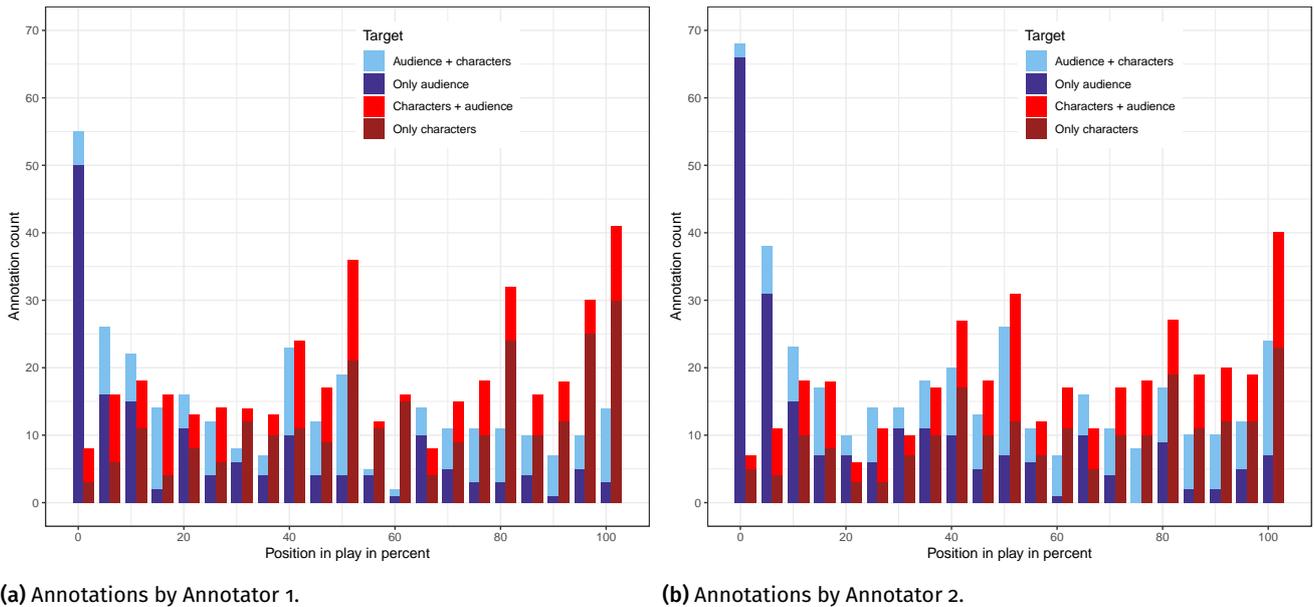
	Annotation 1		Annotation 2	
	Total	Percent	Total	Percent
No. of characters	289	NA	289	NA
No. of characters involved in transfer	134	46.37	127	43.94
... as SOURCE	105	36.33	95	32.87
... as TARGET	118	40.83	111	38.41
Character relays information about themselves	258	0.47	215	0.42
Character receives information about themselves	32	0.06	24	0.05

Table 6: Overview of characters and their involvement in knowledge transfers.

second rank. It is most frequently used for characters that are at first mentioned without name and therefore annotated by a variable that is later unified with their character id. Unsurprisingly, the relations `child_of` and `parent_of` are also frequently used. These mark the importance of the core family for the plot of our selected plays.

When combining annotation 1 and 2, around 50 % of the characters are involved in knowledge transfers, with 38 % being the source and around 43 % being the target of knowledge transfers at least once. In 45 % of the cases, the SOURCE transfers a relation involving themselves. In contrast, in only 5 % of all cases, the TARGET learns about a relation concerning themselves. It is evident that characters possess the most knowledge about their own relations and can therefore pass on this knowledge reliably. For the same reason, learning about one's own family or love relations is rather rare, but might point to especially interesting passages of the plot, as *Udohla* has exemplified. Table 6 shows a detailed breakdown of the numbers per annotator.

Additionally, we also investigate when in a play knowledge transfers happen. Figure 2 shows the number of annotations over the relative position at which they occur and whom they are directed at: other characters in the internal communication system, the audience, or both. The position in this analysis encompasses the entire text, including the dramatis personæ. We bin the number of annotations, so that each bar covers a range of 5 %. Thus, 55 annotations were made by Annotator 1 in the first 5 % of all plays with the audience as the target, 26 annotations were made in the next 5 %, and so on. We can see that the segments right at the beginning and end of a play are the ones with the highest number of annotations. The remaining segments of the plays have a more or



(a) Annotations by Annotator 1.

(b) Annotations by Annotator 2.

Figure 2: Distribution of annotations by Annotator 1 (2a) and Annotator 2 (2b) over the relative positions of the 20 plays. The annotations are separated by the target of the knowledge transfer: (i) Only the audience is the target, (ii) only characters are the target, but not the audience, or (iii) the audience and one or more characters are the target. Note that the latter is displayed twice to get the correct total for both audience and characters.

less similar distribution of annotations with increases in the middle of the plays and in the final quarter. At the beginning of the plays, the majority of information is transferred to the audience (blue bars), while this focus shifts to the characters towards the middle and end of the plays (red bars).

This observation can be explained conclusively as it supports established drama theory. The beginning and end of a play are central places for the transmission of knowledge, both in the internal and the external communication system. “What we understand as the transmission of information at the beginning of a play largely coincides with the classical theoretical concept of the exposition,” acknowledges Manfred Pfister (1988, 86). He goes on to define the exposition as forwarding of information concerning “events and situations from the past that determine the dramatic present”.²⁵ With regard to the audience, the transmission of information in the character’s internal communication system or the *dramatis personæ* fulfils at least two functions: On the one hand, it is intended to instigate the audience’s attention. On the other hand, the audience is provided with the knowledge necessary to understand the subsequent actions (cf. Asmuth 2016, 103–105). As Figure 2 illustrates for our annotated corpus, most of the knowledge about family and love relations that is transferred at the beginning of a play is indeed directed at the audience – oftentimes even solely. Similar to the exposition at the beginning, the resolution at the closure of a play is a common section for transmitting unknown information, e. g., through recognition. In such closed endings, deviations in knowledge between characters and the audience are typically dissolved: “as a result of either intrigue, self-deception or lack of information”, a character or even a group of characters have gotten into trouble. “This situation then culminates in either a happy or

25. The exposition, then, is not necessarily limited to a play’s introduction. Furthermore, not every information that is transmitted early on serves an expository purpose.

a tragic ending, after additional information has been introduced” (Pfister 1988, 95). The values in Figure 2 show a shift of direction towards the end of the annotated plays. About halfway through the plays, the number of annotations directed at characters in the internal communication system increases relatively to those directed at the audience. The knowledge that is transmitted in the resolution is frequently addressed to the plays’ characters. The audience, in turn, already possesses the information necessary to deduce the probable outcome. Thus, the suspense felt by the audience at the end – at least in our corpus – seems to be in respect to *how* an information they possess influences the characters’ actions.

5.2 Tracking Knowledge Flow

We can exemplify these theoretical considerations that relate to the communication system when inspecting the distribution of a specific piece of information as presented in Figure 3. The two visualisations of knowledge transfers in Günderröde’s *Udohla* present an instructive way to display the flow of knowledge within the play. By discussing the possible marriage between the Sultan and Nerissa at the beginning of the play, Sinu, Mangu and the Dervish indirectly pass on their knowledge to the audience. In doing so, the audience is also informed that Nerissa and the Sultan are siblings – at least according to the current beliefs of the present characters. As Figure 3a visualises, it is Nerissa herself that corrects this wrong piece of information for the audience while she is talking to Elpa. From there on the audience has an information advantage over most of the fictional characters. For the other characters, it takes until the middle of the second act, where Mangu receives a letter of the Sultan’s actual sister, to learn of this fact. Udohla, in the meantime, passes along the information that he is Bahadar’s son to Sino, who then passes the knowledge to Mangu between the scenes (see Figure 3b). Mangu in turn tells the sultan. The resolution at the end of the play then brings together the knowledge acquired by the various characters in the course of the play. Nerissa reveals that she is the daughter of Bahadar. She is the last character to learn that Udohla, too, is Bahadar’s child and thus her brother. The visualisations concisely depict the flow of these pieces of information and how they are transmitted from character to character or to the audience.

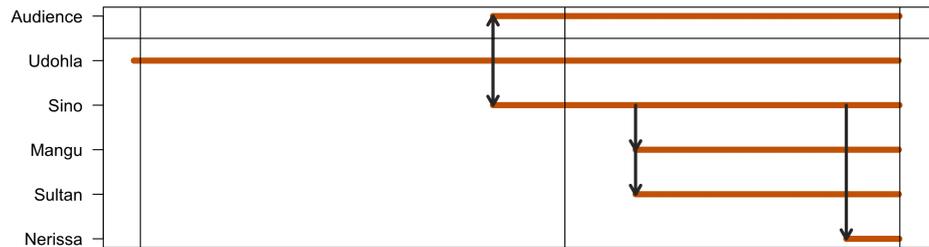
5.3 Networks of Knowledge Transfer

As a third kind of analysis, we use the annotated knowledge transfers to construct character networks. Networks, which are based on the knowledge about family relations and its dissemination in a play, can help to identify key characters that propel the dramatic plot either by gaining new information or by distributing it. In these networks, each node represents a character (or other sources of information such as letters, observations, etc.) and edges between nodes signify that one or more family-related knowledge transfers between two nodes have taken place.²⁶ They can therefore be used to complement the information gathered by established configuration based networks that focus on co-presence of characters. Since there is a SOURCE and a TARGET for each knowledge transfer, the networks are directed. The edges can be weighted with the total number of

26. To compute the knowledge transfer networks, we only focus on the internal communication system of the dramatic characters. Therefore, we omit the audience’s nodes and the dramatis personæ in the networks.



(a) Knowledge flow for the information `!siblings(sultan, nerissa)`.



(b) Knowledge flow for the information `child_of(udohla, "Bahadar")`.

Figure 3: Knowledge flow over the two acts of Günderrode’s *Udohla*.

knowledge transfers that have taken place between two nodes. An example of such a network is shown in Figure 4 for Günderrode’s *Udohla*. The nodes are scaled according to their weighted degree (Barrat et al. 2004), which is a measure that calculates the sum of the weights of all incoming and outgoing edges for each node.

The visualisation in Figure 4 shows Udohla, Sino, who is a Hindu staff member of the Sultan, and the vizier Mangu, to be the central characters of the network according to their weighted degree. At first glance, it might seem surprising that Sino and Mangu are two of the most central characters in the network. For the plot and its resolution, there are more important characters, mostly Nerissa,²⁷ Udohla and the Sultan. How can the central position of Sino and Mangu, i. e., their high weighted degree then be explained? For Sino, there are mainly two reasons. The first reason concerns the intra-fictional progression of the plot. Günderrode conceptualised Sino as Udohla’s only confidant within the Sultan’s palace (cf. Obermeier 1996, 106–107). Both Sino and Udohla are Hindus and they are linked through a mutual close acquaintance. Naturally, then, Sino is the only character in the play Udohla could trust to share his real identity with, which is important for the play’s final resolution, as Sino is able to confirm to the Sultan that Udohla is Bahadar’s son. The second reason is that Sino’s role is used to transmit knowledge from the internal communication system to the audience. Herein, Sino becomes the recipient of new information, while primarily the audience is “the intended receiver of the information given.” (Pfister 1988, 89) To that effect, Mangu takes on a different role in the network. As he receives the letter of the Sultan’s real sister, he is then able to pass on the information that Nerissa is not the Sultan’s sister to other characters of the play. The audience, however, already knows this fact from an earlier

27. As Obermeier (1996, 102–103) states with regard to Nerissa, “her actions alone effect the resolution of the dramatic conflicts.”

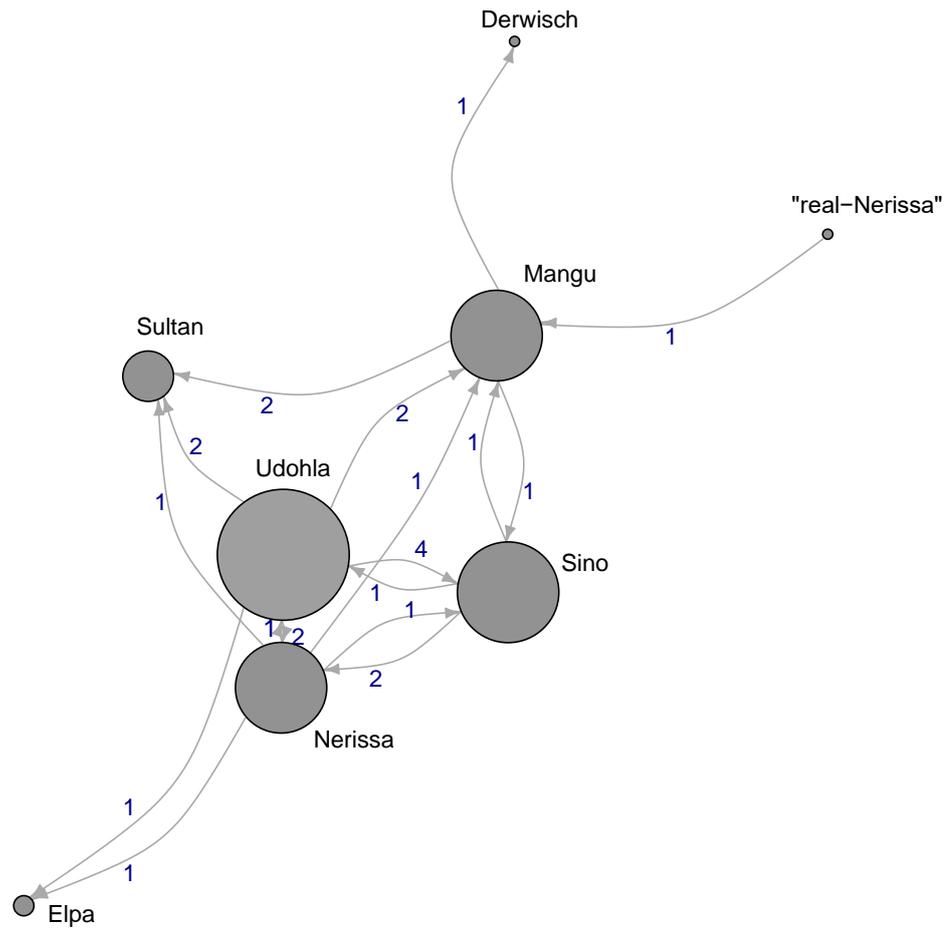


Figure 4: Network of knowledge transfer in Günderrode’s *Udohla*. Based on the consensus version.

conversation of Nerissa and Elpa.

To further track the development of knowledge in the course of *Udohla*, we bin the play’s text into 10 equal-length segments and create a network in each of these segments. In these networks, we calculate in- and out-strength. While the strength metric to scale the nodes in [Figure 4](#) uses both incoming and outgoing edges, in-strength only considers incoming, and out-strength only considers outgoing edges for the calculation. [Figure 5](#) shows cumulative curves for the development of both in-strength and out-strength in *Udohla*. Here, *cumulative* means that the networks of each bin are constructed by taking the annotations of the current bin and all previous bins into account. In this way, we can see which character received and transferred knowledge about family relations at what point in the play. There are some instructive observations that are in need of interpretation. Firstly, Udohla’s high out-strength value is mostly linked to a single scene in the middle of the play, where he introduces himself as Achmed pretending to be the Nawab’s herald. As he passes this false information to five other characters, it has a big impact on his central position in the network. Secondly, the roles of Sino and Mangu in the knowledge transfer network seem to be roughly comparable. Both receive information about family relations that they in turn pass on to other characters. While Sino can be described as a confidant of Udohla, Mangu, being a Muslim, takes on a similar role with regard to the Sultan. All of the important information that the

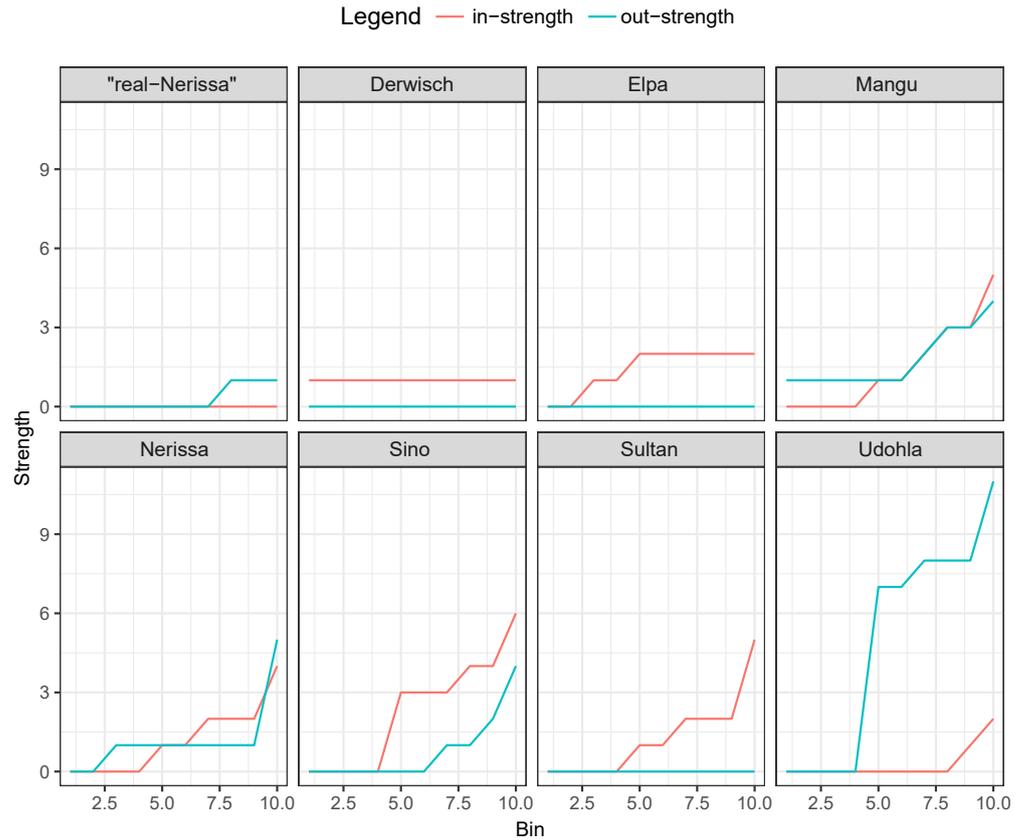


Figure 5: Cumulative in-strength and out-strength in the course of Günderrode's *Udohla* for all the involved entities.

Sultan receives before the final resolution comes from Mangu. Thirdly, the Sultan's role in view of knowledge distribution is strikingly passive. He is only TARGET of knowledge transfers, never the SOURCE. This underlines a different conceptualisation of the Sultan's character. Although he does indeed receive some information in the course of the play, oftentimes he is the last character to be reached by this knowledge. Looking at the play's resolution, this makes sense. Being at the centre of the final recognition scene, the Sultan has to be unaware that Nerissa and Udohla are the children of Bahadar up to this point. Sino and Mangu, on the other hand, accumulate new knowledge throughout the play and serve as middlemen, bridging the knowledge either to the audience or to the main characters.

Following this, we investigate the so-called betweenness centrality (Freeman 1977) of the network. Betweenness centrality measures how often a node is part of a shortest path between two other nodes (cf. Freeman 1977, 37). Since betweenness centrality can be seen as a measure for the flow of communication in a network and how single nodes control the flow of communication, it appears to be especially suited for networks of knowledge transfer. Its "use seems natural in the study of communication networks where the potential for control of communication by individual points may be substantively relevant" (Freeman 1977, 40), as Linton Freeman states in his pioneering study. Figure 6 shows the development of the betweenness centrality for *Udohla*. We can see that Sino and especially Mangu are the characters with the highest betweenness centrality in the play. This further corroborates their role as middlemen in the play.

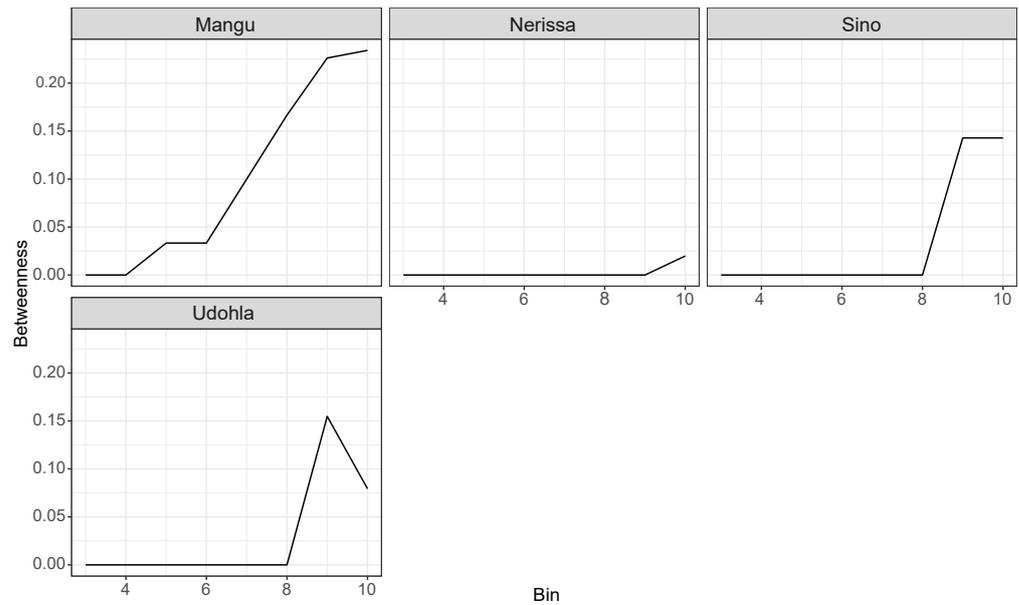


Figure 6: Betweenness centrality in the course of G nderrode’s *Udohla* for all the involved entities. Three characters received a betweenness centrality value of 0 for all positions and were omitted from the graph: Derwisch, Elpa and the Sultan.

Moreover, the visualisation illustrates that in *Udohla*, knowledge transfers responsible for betweenness centrality mostly occur in the second half or even the end of the play. Looking at the structure of a theatre play, this seems conclusive from a conceptual point of view. As a node has to be both TARGET and SOURCE of at least one knowledge transfer to be part of a shortest path, it is not surprising to find this realised only towards the end of *Udohla*. As shown above, the beginning and end of a play are key segments for the transmission of knowledge. In order to have a play’s resolution resulting directly from a recognition scene – as is demonstrated in G nderrode’s *Udohla* – the characters involved must possess a different knowledge base right until that moment.

In summary, the analyses have raised a promising perspective for more extensive investigations on a bigger corpus. Our annotation data can provide insights into different structural principles of German plays. Sino’s and Mangu’s central position in the network and their values for in- and out-strength as well as betweenness centrality further show the potential of our methodological approach. As exemplified in *Udohla*, we can detect characters that have a key role for the flow of knowledge in the course of the play, without being considered main characters themselves. Although our networks are based on the transmission of knowledge about family relations, they depend on co-presence networks. Thus, they can be described as second-order networks. I. e., if in the course of a play two characters are not present on stage together, it is highly unlikely that new information circulates between them. Therefore, we consider a systematic comparison between co-presence networks and knowledge transfer networks as an especially insightful task for future research.

6. Conclusions

In this article, we have presented a composite scheme for the annotation of knowledge transfers about family relations in German plays. As illustrated throughout our article,

annotating these knowledge transfers is a complex task, which gives rise to a number of challenges. Our scheme is based on considerations of drama theory on knowledge distribution. As our results are prospectively also intended to be of relevance for research in traditional literary studies, we have refrained from an operationalisation that overly simplifies concepts in light of computation. Instead, we chose an operationalisation that purposefully connects to terms and concepts of drama theory. As a consequence, the scheme is situated at the intersection between annotation and modelling.

At the same time, this project is (to our knowledge) the first that attempts to measure inter-annotator agreement for such a complex annotation task by employing the metric Gamma. We have discerned a number of intricacies that make the application of Gamma tricky and might be relevant for other annotation projects in computational literary studies. While the ability to provide a custom similarity function makes Gamma versatile, this also requires us to make a high number of design decisions that influence the results and decrease comparability with other applications of Gamma. Conceptually, the definition of what we want to consider a positional and/or categorical agreement is not always straightforward because of the (sometimes) vague nature of the target phenomenon, the compositionality of the annotation labels, and dependencies between its components.

As our preliminary analyses have shown, a systematic annotation of knowledge transfers about family relations allows for investigations that go beyond structural features of the play's surface. Herein, we made use of our annotation data to propose an extension to the widely utilised co-presence networks. In specifying the edges as a directed knowledge transmission, networks can be interpreted in light of more tailored research questions as we have hinted at with Günderröde's *Udohla*. The analyses have also revealed clear perspectives for larger corpus studies. This gives rise to future questions concerning literary history. Do patterns of family related knowledge distribution emerge for different dramatic genres? Is it possible to characterise the scenes where changes of knowledge occur in more detail? How many characters are on stage in these scenes? How many of them are actively involved in passing on knowledge? What kind of characters do pass on the knowledge?

Our future work mostly focuses on two aspects. Firstly, we are currently implementing a system to automatically infer all deducible family relations from our annotations. As the annotations only cover the transmission of new information from one character to another (or to the audience), this inference system is needed to have a full account of what all characters and the audience know at all times during the drama. Having this knowledge base would both benefit the measuring of the IAA – as it would solve certain problems such as using different predicates for the same relation – and the subsequent analysis. Secondly, we are working on automating certain aspects of the annotation process by creating transformer-based machine learning models which learn to predict the positions in a text where knowledge is transferred and the type of family or love relation that is transferred. Applying these models on new data will facilitate the annotation of new texts. Evaluating the performance of the models on existing data can give additional insights into the complexity of the annotation task.

7. Data Availability

Data can be found here: <https://github.com/quadrama/jcls2022>

8. Software Availability

Software can be found here: <https://github.com/quadrama/jcls2022>

9. Acknowledgements

The research in this article has been conducted in the Q:TRACK project (<https://quadrama.github.io/index.en>), which is part of the priority programme SPP 2207 *Computational Literary Studies* and funded by the German Research Foundation (DFG). We thank Jonas Hirner and Christian Lantzinger for their annotations and Erik Ketzan, Trinity College Dublin, for proofreading.

10. Author Contributions

Melanie Andresen: Annotation Supervision, Guideline Development, Inter-Annotator Agreement, Writing

Benjamin Krautter: Theoretical Framework, Analysis and Interpretation, Writing

Janis Pagel: Corpus Statistics, Network Analysis, Writing

Nils Reiter: Methodology, Inter-Annotator Agreement, Writing

References

- Anderson, Maxwell (1965). "The Essence of Tragedy". In: *Aristotle's "Poetics" and English Literature*. Ed. by Elder Olson, 114–121.
- Anz, Thomas (2007). "Spannung". In: *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*. Ed. by Jan-Dirk Müller, Georg Braungart, Harald Fricke, Klaus Grubmüller, Friedrich Vollhardt, and Klaus Weimar. Vol. 3. De Gruyter, 464–467.
- Aristotle (1995). "Poetics". In: *Aristotle: Poetics*. Ed. by Stephen Halliwell. Harvard University Press, 27–141.
- Artstein, Ron and Massimo Poesio (2008). "Inter-Coder Agreement for Computational Linguistics". In: *Computational Linguistics* 34 (4), 555–596. [10.1162/coli.07-034-R2](https://doi.org/10.1162/coli.07-034-R2).
- Asmuth, Bernhard (2016). *Einführung in die Dramenanalyse*. 8th updated and expanded ed. J.B. Metzler.
- Barrat, A., M. Barthélemy, R. Pastor-Satorras, and A. Vespignani (2004). "The architecture of complex weighted networks". In: *Proceedings of the National Academy of Sciences* 101 (11), 3747–3752. [10.1073/pnas.0400087101](https://doi.org/10.1073/pnas.0400087101).
- Cave, Terence (1988). *Recognitions. A Study in Poetics*. Clarendon Press.

- Destrée, Pierre (2020). "Family Bounds, Political Community, and Tragic *Pathos*". In: *The Poetics in its Aristotelian Context*. Ed. by Pierre Destrée, Malcolm Heath, and Dana L. Munteanu. Routledge, 113–128.
- DiYanni, Robert (2000). *Drama. An Introduction*. McGraw-Hill.
- Dubbels, Elke (2012). "Zur Dynamik von Gerüchten bei Heinrich von Kleist". In: *Zeitschrift für deutsche Philologie* 131 (2), 191–210. [10.37307/j.1868-7806.2012.02.03](https://doi.org/10.37307/j.1868-7806.2012.02.03).
- Dutant, Julien (2015). "The Legend of the Justified True Belief Analysis". In: *Philosophical Perspectives* 29 (1), 95–145. [10.1111/phpe.12061](https://doi.org/10.1111/phpe.12061).
- Engelstein, Stefanie (2004). "Sibling Incest and Cultural Voyeurism in Günderode's "Udohla" and Thomas Mann's "Wälsungenblut"". In: *The German Quarterly* 77 (3), 278–299.
- Evans, Bertrand (1960). *Shakespeare's Comedies*. Clarendon Press.
- Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora – Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor". In: *DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts*, 194–197. [10.5281/zenodo.2596095](https://doi.org/10.5281/zenodo.2596095).
- Fleiss, Joseph L. (1971). "Measuring nominal scale agreement among many raters". In: *Psychological Bulletin* 76 (5), 420–428. [10.1037/h0031619](https://doi.org/10.1037/h0031619).
- Freeman, Linton C. (1977). "A Set of Measures of Centrality Based on Betweenness". In: *Sociometry* 40 (1), 35–41. [10.2307/3033543](https://doi.org/10.2307/3033543).
- Gettier, Edmund L. (1963). "Is Justified True Belief Knowledge?" In: *Analysis* 23 (6), 121–123.
- Günderode, Karoline von (1990). "Udohla". In: *Karoline von Günderode: Sämtliche Werke und ausgewählte Studien*. Ed. by Walter Morgenthaler. Vol. 1. Stroemfeld/Roter Stern, 203–231.
- Ichikawa, Jonathan Jenkins and Matthias Steup (2018). "The Analysis of Knowledge". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/> (visited on 10/24/2022).
- Jaccard, Paul (Feb. 1912). "The Distribution of the Flora in the Alpine Zone". In: *The New Phytologist* 11 (2). [10.1111/j.1469-8137.1912.tb05611.x](https://doi.org/10.1111/j.1469-8137.1912.tb05611.x).
- Jeßing, Benedikt (2015). *Dramenanalyse: Eine Einführung*. Erich Schmidt Verlag.
- Kablitz, Andreas (1998). "Wiedererkennung: Zur Funktion der Anagnorisis in der klassischen französischen Tragödie (Corneille: *Œdipe – Racine: Iphigénie en Aulide*)". In: *Erkennen und Erinnern in Kunst und Literatur*. Ed. by Dietmar Peil, Michael Schilling, and Peter Strohschneider. Max Niemeyer Verlag, 455–486.
- Krautter, Benjamin and Marcus Willand (2021). "Vermessene Figuren: Karl und Franz Moor im quantitativen Vergleich". In: *Schillers Feste der Rhetorik*. Ed. by Peter-André Alt and Stefanie Hundehage. De Gruyter, 107–138.
- Krippendorff, Klaus (2004). *Content Analysis: An Introduction to its Methodology*. 2nd ed. Sage.
- Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33 (1), 159–174.
- Licher, Lucia Maria (1996). *Mein Leben in einer bleibenden Form aussprechen. Umriss einer Ästhetik im Werk Karoline von Günderodes (1780–1806)*. Universitätsverlag Winter.

- Lipinski, Silke (2011). "Udohla – Plattform für Karoline von Günderrodes philosophische Gedanken". In: *New German Review* 24 (1), 113–122.
- Marcus, Solomon (1973 [1970]). *Mathematische Poetik*. Trans. by Edith Mándroiú. Athenäum.
- Mathet, Yann, Antoine Widlöcher, and Jean-Philippe Métivier (2015). "The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment". In: *Computational Linguistics* 41 (3), 437–479. [10.1162/COLI_a_00227](https://doi.org/10.1162/COLI_a_00227).
- Moretti, Franco (2011). "Network Theory, Plot Analysis". In: *Pamphlets of the Stanford Literary Lab* 2. <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> (visited on 11/08/2019).
- Mueller, Martin (2012). *Scalable Reading*. <https://web.archive.org/web/20211201185120/https://sites.northwestern.edu/scalablereading/2020/04/26/scalable-reading/> (visited on 10/24/2022).
- Obermeier, Karin (1996). " "Ach diese Rolle wird mir allzu schwer". Gender and Cultural Identity in Karoline von Günderrode's Drama *Udohla*". In: *Thalia's Daughters. German Women Dramatists from the Eighteenth Century to the Present*. Ed. by Susan L. Cocalis and Ferrel Rose. Francke, 99–114.
- Pagel, Janis, Nils Reiter, Ina Rösiger, and Sarah Schulz (2020). "Annotation als flexibel einsetzbare Methode". In: *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der Creta-Werkstatt*. Ed. by Nils Reiter, Axel Pichler, and Jonas Kuhn. De Gruyter, 125–141. [10.1515/9783110693973-006](https://doi.org/10.1515/9783110693973-006).
- Pfister, Manfred (1988). *The Theory and Analysis of Drama*. Trans. by John Halliday. Cambridge University Press.
- Pichler, Axel and Nils Reiter (2020). "Reflektierte Textanalyse". In: *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der Creta-Werkstatt*. Ed. by Nils Reiter, Axel Pichler, and Jonas Kuhn. De Gruyter, 43–59. [10.1515/9783110693973-003](https://doi.org/10.1515/9783110693973-003).
- (2021). "Zur Operationalisierung literaturwissenschaftlicher Begriffe in der algorithmischen Textanalyse. Eine Annäherung über Norbert Altenhofers hermeneutische Modellinterpretation von Kleists *Das Erdbeben in Chili*". In: *Journal of Literary Theory* 15 (1–2), 1–29. [10.1515/jlt-2021-2008](https://doi.org/10.1515/jlt-2021-2008).
- Pollock, John L. and Joseph Cruz (1999). *Contemporary Theories of Knowledge*. 2nd ed. Rowman & Littlefield.
- Reiter, Nils (2018). "CorefAnnotator – A New Annotation Tool for Entity References". In: *EADH 2018*. <https://eadh2018.exordo.com/programme/presentation/118> (visited on 10/24/2022).
- Reiter, Nils and Leonard Konle (2022). "Messverfahren zum Inter-annotator-agreement (IAA): Eine Übersicht". In: *DARIAH-DE Working Papers* 44. [10.47952/gro-publ-103](https://doi.org/10.47952/gro-publ-103).
- Reiter, Nils and Marcus Willand (2019). "Surveying Shakespeare's Impact on German Drama: Taking a Computational Approach to an Epoch". In: *Anglo-German Dramatic and Poetic Cultures: New Perspectives on Exchange in the Sattelzeit*. Ed. by Michael Wood and Sandro Jung. Lehigh University Press, 117–143.
- Schadewaldt, Wolfgang (1955). "Furcht und Mitleid? Zur Deutung des Aristotelischen Tragödiensatzes". In: *Hermes* 83 (2), 129–171.
- Schmitt, Arbogast (2008). "Kommentar". In: *Aristoteles: Werke*. Ed. by Hellmut Flashar. Vol. 5, 193–741.
- Trilcke, Peer (2013). "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft". In: *Empirie in der Literaturwissenschaft*. Ed. by Philip Ajouri, Katja Mellmann, and Christoph Rauen. Mentis, 201–247.

- Trilcke, Peer, Frank Fischer, and Dario Kampkaspar (2015). "Digital Network Analysis of Dramatic Texts". In: *Digital Humanities 2015: Global Digital Humanities. Book of Abstracts*. [10.5281/zenodo.3627711](https://doi.org/10.5281/zenodo.3627711).
- Yarkho, Boris I. (2019 [1935–1938]). "Speech Distribution in Five-Act Tragedies (A Question of Classicism and Romanticism)". In: *Journal of Literary Theory* 13 (1), 13–76. [10.1515/jlt-2019-0002](https://doi.org/10.1515/jlt-2019-0002).