Schriften des Instituts für Dokumentologie und Editorik — Band 16

Digitale Edition in Österreich
Digital Scholarly Edition in Austria

---

herausgegeben von | edited by

Roman Bleier, Helmut W. Klug

2023

BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de/ abrufbar.

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 29. April 2023.

2023

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-743-102-842
Einbandgestaltung: Stefan Dumont; Coverbild: wurde von Roman Bleier und Helmut Klug für ein KONDE-Poster (DHa 2017) erstellt
Satz: Roman Bleier und LuaTEX

A Linked Data Vocabulary for Intertextuality in Literary Studies, with some Considerations Regarding Digital Editions

Bernhard Oberreither

Abstract

Editing a text often includes pointing out its manifold connections to other texts—be it cultural contexts, sources, influences, or texts evoked by quotations and allusions—, all of which can be subsumed under a broad notion of 'intertextuality'. In light of the opportunities provided by the Semantic Web, this paper presents an OWL 2 ontology for the modelling of these intertextual relations—with respect to the fact that the identification of said relations is in itself an act of interpretation that must be reflected upon as such.

Zusammenfassung

Einen Text zu edieren beinhaltet nicht zuletzt, seine vielfältigen Verbindungen zu anderen Texten aufzuzeigen, seien das kulturelle Kontexte, Einflüsse, zitatweise oder in Anspielungen aufgerufene Texte etc.; alles das fällt unter den (weiter gefassten) Begriff der Intertextualität. Angesichts der Möglichkeiten des Semantic Web wird hier eine OWL 2-Ontologie für die Modellierung dieser intertextuellen Verknüpfungen vorgestellt, die dabei die Tatsache berücksichtigt, dass jede Feststellung von Intertextualität zugleich einen interpretativen Akt darstellt, der als solcher reflektiert werden muss.

1 Edition, commentary, and interpretation

Every edition is an interpretation. Every conclusion drawn by the editor based on his or her material can be seen as an interpretive act. This is perhaps most obvious regarding the commentary. As a consequence, over the past decades, very different paths have been proposed for the scholarly edition: some argue that the ideal editor has to exercise interpretive abstinence wherever she can (e.g., Martens 1993), restraining herself especially when it comes to adding a commentary to the edited text. Others suggest that commentary is not to be avoided, even more so: it is not to be restrained to selective (for example, biographical, historical) data on individual text passages.

Digitale Edition in Österreich. Hrsg. Roman Bleier und Helmut W. Klug. Schriften des Instituts für Dokumentologie und Editorik 16. Books on Demand, 2023, 69–87.

Instead, the commentary should also discuss semantic and poetic dimensions of the text as a whole, which cannot be done without proposing hypotheses (e.g., Windfuhr 1991, 173f.)—thus engaging in the business of interpretation. Some have drawn radical conclusions, arguing for a section titled 'interpretation' as an equal part of a scholarly edition, pointing out that what is usually called 'commentary' can only to a very small degree be clearly separated from an interpretation (Stüben 1993).

While there seems to be no clear distinction between these two, in order to determine where commentary/interpretation is appropriate one could refer to a traditional differentiation: the differentiation between 'primary obscurity', where obscurity is a function of, for instance, poetics and interpretation should be left to the reader, and 'secondary obscurity', where meaning has become obscure due to historical and/or cultural distance and the reader might be misled without the help of the editor (Fuhrmann 1985; Martens 1993, 40f.). In practice, this opens a grey area—and by doing so, transfers the problem of blurred boundaries from the question of commentary and interpretation to yet another unclear distinction.

Commentary cannot be kept free from interpretation; this is also—perhaps especially—true for a specific sub-domain of commentary: accounting for the intertextual relations of the text in question, traditionally under the notion of 'sources'. This domain has been perceived as "the one closest to the actual task of the editor, publishing a text and its variants" and thus has been used to argue in favour of the justification of the commentary section (Groddeck 1993, 2; transl. B.O.[1]). Still, even the apparently simple act of pointing out a source can interfere with underlying structures of meaning and, consequently, effect the text's reception (cf. Groddeck 1993, Kanzog 1999).

Linking a text to its sources has always had certain theoretical implications: the term 'source' can have a broader or a narrower definition (a narrower distinguishing it from meanings like 'influence', 'reminiscence', 'allusion'; Woesler 1997), but the theory behind it usually includes, broadly speaking, an author reading a text, which then in some way affects her writing of a text. The term 'source' implies some kind of chronology, a positivistic cause-and-effect-chain, and maybe an insight into the author's intentions. However, concepts of text-text-relations such as these were complicated by the emergence of intertextuality theories. The relevance of other, more distinct types of intertextual relations was pointed out, and the editorial preference of explicit quotes to implicit ones was questioned (Zittel 1997).[2]

However, the more radical theories of intertextuality, especially the poststructuralist (where the term—not the notion or investigation—of intertextuality was coined), were widely considered impractical and bearing the danger of arbitrariness; among

---

[1] "[D]ie Aufgabe, Zitate im Text nachzuweisen, liegt der eigentlichen Aufgabe des Editors, den Text mit seinen Varianten herauszugeben, am nächsten." (Groddeck 1993, 2).

[2] Groddeck aims to 'infect' scholarly editing with the 'virus of intertextuality theory' (Groddeck 1993, 2).

editors, more restrictive classification systems for intertextual relations were deemed applicable, to make explicit the exact nature of any text-text-link (Pfister/Broich and the clearly structured catalogue of the different aspects of an intertextual relationship in Pfister 1985 seemingly met the editors' taste; see Groddeck 1993, Kanzog 1999, Zittel 1997).

It is probably safe to say that it was not only the danger of arbitrariness that limited the amount (and type) of intertextual relations accounted for in an edition but also issues closely related to the print medium: readability and limited space. Thus, in the eighties and early nineties, hypertext-theory and its realisation in HTML were welcomed as the solution for editorial problems such as these. Among other things (like the direct linking of the text and commentary/apparatus, interactive user interfaces and the possibility of multiple perspectives on a text), the representation of intertextual relations and the integration of unlimited amounts of source material were expected to be facilitated and enabled by the digital medium (Sahle 2013, 41–52). Gunter Martens's concept of the 'intertextual commentary' from 1993 already pointed towards what was later discussed under the label 'edition as archive', where an edition formerly referencing external archive material now includes these materials (Martens 1993, Sahle 2007). Martens used the term 'intertextuality' to elegantly evade the problem of interpretation. In his concept of an 'intertextual commentary,' intertexts of any kind and the edited text would be published side by side, without specification of the respective relation: "The processing of these materials is not to be forced upon the reader. The drawing of conclusions and integrating them in a reading of the poetic text is left to the recipient's own interpretive operation" (Martens 1993, 49; transl. B.O.[3]).

However, what looks like a new technology—the hypertext—and a methodological concept—the 'intertextual commentary'—coming together to solve the problem leaves one crucial point untouched: Martens's proposal helps avoid explicit interpretation, but not only leaves the interpretation to the reader, but also has its own theoretical implications. Unless the 'intertextual commentary' consists of an unfiltered and unstructured amount of texts (which would render it useless for its intended purpose, but would make for a useful tool of its own; Baßler/Karczewski 2009), it has to undergo processes of selection and combination, processes that contain an interpretive aspect and certainly draw from theoretical prerequisites.

Wherever one stands in this discussion about which intertextual relations to account for and which materials to include—it is fair to say that as soon as an intertextual relation is pointed out, some kind of theoretical framework plays a role and should be made explicit to uphold scholarly transparency.

---

[3] "Die Verarbeitung solcher Materialien ist dabei dem Leser nicht vorgegeben. Schlüsse daraus zu ziehen und in die Deutung des poetischen Textes einzubringen, ist der eigenen interpretierenden Arbeit des Rezipienten überlassen." (Martens 1993, 49).

2 Building and transcending the 'data silo'

Let us assume a basic digital edition concerned with a text's intertexts (for instance, the author's sources): it would usually include, besides the edited text, a document containing said sources/intertexts (maybe as a list of bibliographical references). These two components can have the form of two TEI-XML documents. Specifying theoretical and methodological implications in the most general way comes down to specifying the nature of the link between a passage of the edited text in the one document and the data representing its intertext in the other. For this purpose, TEI provides several options (TEI-Consortium 2020)—an obvious one would be to add the attributes `@type` and `@subtype` to a `<quote>` element. The explanatory power of these possibilities is, however, quite limited: attribute values can be used in a self-explanatory way, leading to a limitation to common terms without regard for the theoretical background. Otherwise, attribute values can point to a more sophisticated classification of intertextual relations that then needs to be underpinned in a statement in natural language somewhere in the document's metadata or a website's editorial note.

This is one example of why 'traditional' digital editions tend to fall under the verdict of being 'data silos': of not representing the published statements (about the edited text, its sources, etc.) in an interoperable or—ideally—a machine-readable way. In this respect, such an edition is no different from a printed one (cf. for example, van Zundert 2016, 101). The intertextual relations accounted for in the edition are neither classified as such nor differentiated in any way that, to a non-human recipient, would not appear as white noise. Furthermore, the notion of the data silo also includes what I would like to call a 'theory silo' since without classification an assessment of theoretical implications is impossible.

There lies a huge potential for interconnectivity in an edition's list of intertexts. This is the place where an edition opens itself to literary or cultural discourse. In many cases, the referenced texts might already exist in a digitized form somewhere on the Internet. Linked Data is an obvious choice to represent these intertextual relations.[4] If an edition exposes its intertexts through Linked Data, short- or medium-term benefits would be interoperability, data-retrieval and -visualization. A visionary long-term perspective would include a decentralized database of intertextual relations throughout numerous online editions and archives; a Linked-Data-Cloud of literature and literary studies.

[4] Early on, a conceptual analogy between the RDF-triple and intertextual relation was noticed—both consist of two elements linked by a specified relation (Tennis 2004). Fabio Ciotti has pointed out that enrichment of data based on Linked Data vocabularies "has similarities with the humanities tradition of annotation and comment" (Ciotti 2014).

Projects and concepts in the digital humanities that tackle the data silo problem build on Linked Data technologies. The *Dante Sources Project*, for example, provides RDF-data on Dante's primary sources, cited authors and more via a SPARQL-endpoint (Bartalesi and Meghini 2017). Interoperability, however, is just the most obvious benefit: as Fabio Ciotti has pointed out, terms and concepts in humanities and cultural studies are often vague and used ambiguously, relying on the reader's prior knowledge and intellectual ability to infer intended meaning. In the course of semantic modelling and formalising, this kind of implicit knowledge is made explicit (Ciotti 2015); the vocabulary to do so is, as part of the technical prerequisites of Linked Data, accessible and provides context for any term or concept inside a conceptual hierarchy. In addition, ontologies in the Semantic Web are meant to represent only a segment of the world as perceived from one specific epistemological point of view (and explicitly so).[5] Every 'thing' in the Semantic Web, every URI, can be subject to numerous classifications based on various ontologies, giving way to a coexistence of epistemologies.

There are a couple of preliminary studies concerned with the utilization of Linked Data in literary studies—mostly of a merely recommending nature. In 2010, Adrian Pohl in a blog-entry emphasized the potential of bibliographic metadata in Linked Data formats, suggesting 'participative catalogues,' where implicit relations between texts could, by way of crowdsourcing, be added to the data (Pohl 2010). In a 2011 article, Sarah Bartlett and Bill Hughes illustrated a possible Linked Data approach to intertextuality based, like Pohl's proposal, on Gérard Genette's terminology (Genette 1997). Charlotte Brontë's *Jane Eyre* is used as a test case in a dry-run-modelling of the text's intertextual (and other) relations (Bartlett and Hughes 2011). In a paper from 2014, Pauline Rafferty evaluated available options for the representation of intertextuality in Linked Data. This possibility is provided by web platforms like 'LibraryThing,' albeit, according to Rafferty, with some conceptual shortcomings. She has suggested a typology that would expand the 17 types of intertextuality provided there (Rafferty 2014).

There are also projects concerned with intertextuality that already implement Linked Data technology. One of them is the *Dante Sources Project*, which even provides a vocabulary for this purpose (and was awarded a DH Award in 2015 in the category 'Best DH Tool or Suite of Tools'). It allows for searching a number of Dante's works as well as their sources from many different angles, starting from Dante's work, cited literature and authors, thematic areas and so on. The ontology created for these intertextual relations is a tailor-made solution strictly for the project's own means. It provides three distinct types of intertextual relation or citation: `citesAsExplicitCitation`, `citesAsStrictCitation`, and `citesAsGeneralCitation`. The first is a citation made

[5] Elsewhere, I called it the "immanent constructivism" of the Semantic Web; proposals for a better wording are still welcome (Oberreither 2019).

explicit by the author himself, the second is a citation indicated by a scholar, the third is a reference to a concept instead of a specific source (Bartalesi and Meghini 2017). While this ontology fits the project's research question, its domain regarding intertextuality is limited to these three types of intertextual relations, reducing its reusability to projects with matching research interests. This limitation is also an issue in approaches suggested by Rafferty or Bartlett/Hughes.

3 Opportunities and desiderata

Thinking about intertextuality in digital literary studies (and here especially editing), we have already come across a couple of things to consider when going forward in constructing a Linked Data vocabulary for this purpose. One aspect worth pointing out is that the concepts and projects referred to above are based on a certain notion of intertextuality that limits its scope to a fixed set of relation types, or to relations inside the realm of literature. If interpreted in a broader way, however, the notion of intertextuality covers not only a possibly infinite number of types of said relations but also the relations between different types of texts. These could concern literary and non-literary texts (for example, scholarly) as well as relations across this divide, when, for instance, a scholarly text analyses a literary text. Intertextuality can also be determined between a text and a text corpus (when questions of classification regarding literary epochs, movements, schools, discourses, etc. arise) as well as between a text and abstract concepts like rhetorical devices, motifs, etc. Thus, modelling intertextuality means modelling a wide variety of scholarly activities. This aspect adds to the relevance of the domain as well as to the scope of an ontology covering it. Before proceeding to the architecture of a vocabulary that would take all this into account, let us sum up the desiderata we have come across so far, and elaborate:

Interoperability: Linked Data is a promising strategy to address the data silo problem. Ontologies, as the cornerstone of interoperability, provide the machine-readable vocabulary that allows for merging different data sets. Interoperability, in this respect, relies on standardization, and it is exactly this aspect that has caught the critical attention of humanists. Standardization opposes individuality of approach; standardization based on the epistemology of one field of research may not fit the needs of another. While the main standard in digital text editing—TEI-XML—is not a standard, since in practice it is open to adaptation and one and any problem can be approached in many different ways, machine-readability calls for stronger restrictions. There are several vocabularies in fields neighbouring literary studies that can serve as points of contact or reference. Alignment to existing standards is a key aspect in Linked Data, but these standards have to be evaluated with respect to their methodological/epistemological and practical implications.

Domain specificity: In respect to the applicability to the domain of literary studies (and probably other research areas inside the humanities as well), any existing Linked-Data-vocabulary—if an alignment makes sense—will have to be extended for the humanities scholars' use. Since links between, to and from literary texts concern more than just the wording of a text, classes and relations tailored for the domain of literary studies will have to address the seemingly simple, yet profound question: what is it that we are actually talking about when we talk about (literary) texts?

Universality: Any vocabulary for the purpose of representing intertextuality in literary studies has to include as wide a variety of methodologies as possible, in order to not involuntarily exclude relevant ones. This means finding a modelling practice that allows for unlimited types of intertextuality; this also means looking for a common denominator of relevant methodological frameworks, as well as sorting out aspects that, if implemented in the modelling, would limit its reach.

These three sections already hint at the design of an ontology covering the domain in question; at the same time, addressing them sometimes means weighing one against the other two (maybe every vocabulary can be characterized by a specific balance of these aspects). Accordingly, the following subchapters give an insight into the considerations leading to the architecture of the proposed ontology, and possibly some points of thought for other, related ventures. Afterwards, an important final aspect will be addressed: the question of practicability, with special attention to the applicability for scholarly editions.

3.1 Interoperability

Looking for a vocabulary to model statements about intertextuality, one soon stumbles upon CiTO, the *Citation Typing Ontology*. This ontology is part of the SPAR ontology suite, a collection of ontologies for the purpose of enriching publications with Semantic Web-metadata (Peroni/Shotton 2018). CiTO provides a list of no less than 43 types of citation, ranging from `agrees with` to `uses method in`, including `critiques`, `discusses`, `parodies`, `plagiarizes` and so forth. The ontology, as these examples show, is designed to represent citations inside academic discourse. Since citation practices are very much standardized in academic practice, the extensive list provided likely covers its domain to a large extent. However, it is still a closed list, and one that is, by design, not meant to cover the myriad ways of intertextual reference found in literature, nor the types of intertextuality conceptualised in literary theory. CiTO's citation types are modelled as object properties, linking a subject and an object, such as a citing and a cited, a plagiarizing and the plagiarized text (see Figure 1 [a]). As can be seen from these examples, relations modelled with CiTO are unidirectional. This makes sense if applied to a scholarly corpus, where one wants to model chains of citations running through a text corpus as chronological chains of cause and effect. In

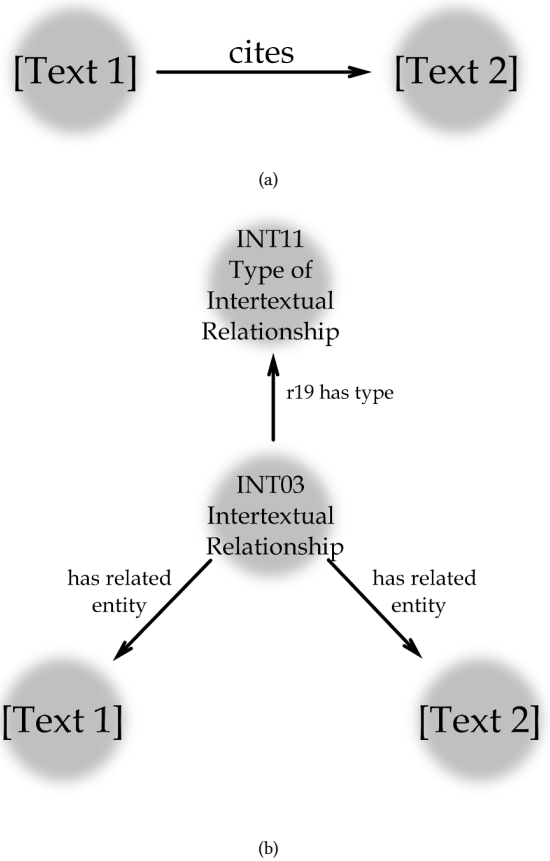

Figure 1: Intertextual relations modelled as object properties or as instances with an attached type.

literary studies, however, bi- and a-directional relations are also quite common. As a most prominent example for bi-directional relations, one can point to Julia Kristeva's notion of a textual universe where 'meaning' in texts is the product of mutual "absorption and transformation" (Kristeva [1967]). An a-directional intertextual relation might be found in relations that entirely depend on a scholar's reading (Bronfen's 'cross mapping' could be interpreted as an example for a methodology of this kind; Bronfen 2002).

As a consequence of the above, the proposed vocabulary first of all provides a class, not object properties, for the intertextual relation (thus reifying the rela-

tion;[6] see Figure 1 [b]). To specify the relation, the user creates an instance of `Type of Intertextual Relationship`, linking it to the relation. It should be mentioned at this point that CiTOs object properties—the list of 43 types of citation—can be re-introduced here as instances of this class. Secondly, the object properties linking the relation to the texts constituting the (at least) two anchor points of the relation do not specify its direction; this is done only in the respective sub-properties.

An important aspect is the question of what an intertextual relation relates to, meaning: where to take the classes and properties to model the instances in question (such as texts). Several vocabularies for this purpose, built for the representation of bibliographic metadata in a library context, already exist. The most elaborate one is FRBRoo (Bekiari et al. 2016), which offers classes for texts on every ontological level, from the `Work` as the most abstract idea and the `Expression` as a work in a structured, yet immaterial form, down to its material `Manifestation`.[7]

FRBRoo is aligned to CIDOC-CRM (ICOM/CRM Special Interest Group 2020): its core classes like `Work` or `Expression`, but also the dynamic ones like `Work Conception` or `Expression Creation` are subclasses of respective classes from the CRM. However, the ontology proposed is linked to this model not only indirectly via FRBRoo but also directly. This is due to the broad domain of CIDOC-CRM: it covers the domain of cultural heritage from a museologist's, archivist's or archaeologist's perspective and offers, at a high hierarchical level, a class for `Conceptual Objects` (as opposed to `Physical Things`; s. Le Boeuf et al. 2019). The relevance of this for our case is beyond question, whenever we talk about abstract concepts like textual features or intertextual relations.

3.2 Domain specificity

A characteristic of literature is the eventual indecisiveness of meaning; this includes the 'meaning' of a quotation. Thus, a characteristic of literary and cultural studies is the more or less general agreement that statements in these fields of research are 'just' interpretations, intrinsically tied to their epistemology, their methodology and

---

[6] As part of its alignment to the Open Annotation Standard, this was also implemented in CiTO (Shotton 2013).

[7] One of its most useful classes in terms of bibliographic information certainly is the `Publication Expression`—the text in an immaterial form, yet so specific as to refer to a distinct edition. This makes for a text with year and place of publication as cited in any bibliography. If I am not mistaken, this train of thought leads to another important notion of text, which however is maybe not supported by FRBRoo: a text that is part of a larger text, yet has all the bibliographical information linking it not only to the larger text but also to time and place of publication. This class is needed when modelling a research paper published in a journal, for instance. A `Publication Expression` can, according to domain/range of the relevant property (P165), incorporate another `Publication Expression`, but the examples given for this class only apply to complete publications, not their parts (like a text in an anthology or a periodical).

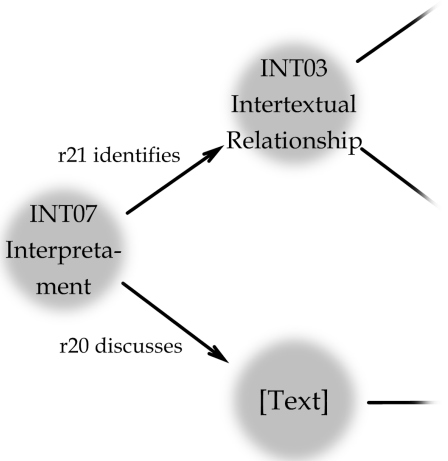

Figure 2: The Interpretament as a means to model the scholarly act.

theoretical background. The humanities have a strong tendency towards reflection on the question of perspective. Thus, a certain feature should not be directly tied to an investigated object. Instead, an instance is needed that accounts for the fact that every attribution is an interpretation. For this instance, the proposed vocabulary provides the class `Interpretament`. Instances of this class serve as the entities to which the identification of features and intertextual relations as well as the more general notion of 'discussing' (a text, a text passage etc.) are tied (see Figure 2). They introduce the meta-level necessary for the discussion of methods and theoretical fundamentals and may be tied to a research output (like a paper, or an RDF document) that they stem from.[8] By making `Interpretaments` explicit, sources and other research findings such as supporting evidence can be linked to them.

The next step is to ask what the `Interpretament` identifies when it identifies a feature of a text. The feature? Hardly ever. Usually, what is being identified is the specific form in which a textual phenomenon resembles other phenomena in other

[8] CIDOCinf, an extension of CIDOC-CRM, models the mental process (the 'inference making') that leads to academic output (Stead et al. 2018); it offers the means to capture argumentation within the descriptive and empirical sciences historically, and thus scientific progress. However, the `Interpretament` excludes the temporal facet represented in the event of inference making; it is rather a representation of its textual traces (where historical markers are only in the respective text's metadata).

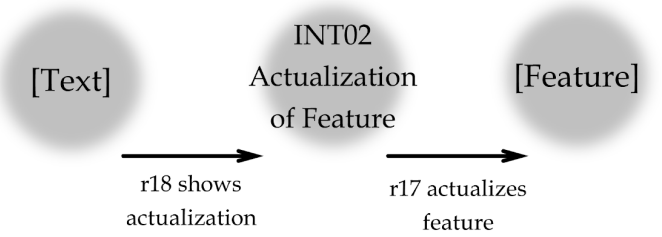

Figure 3: The Actualization of Feature as an intermediary between text and feature.

texts, an abstraction of which is called a textual feature. In the case of the phenomenon in a specific text, we do not talk about features but about the *actualization* of a feature in a text. Thus, a text can never 'have' a feature, it can show its actualization, which constitutes an intermediary between the text and the feature (see Figure 3). This also makes sense for modelling reasons: this way, one can model the common view that the same feature—like a 'Faustian character'—can appear in different texts, and different forms (or actualizations). The class `Actualization` of `Feature` is the link between the text and the feature in question. Also—as is its main purpose here—the class is linked to the intertextual relation, in the many cases where intertextuality does not mean word-by-word-quotation, but instead refers to reoccurring themes, motifs, rhetoric, characters, etc. For example, this allows for the statement of the fact that a fictional text derives one of its characters (maybe a Hanswurst-, Faust- or Sherlock-character) from another text or from some literary tradition.[9]

[9] CIDOC-CRM provides the event `E13 Attribute Assignment`, which in a way merges the classes `Interpretament` and `Actualization` of `Feature`. However, following this model would be inadequate as we cannot reduce interpretaments to events in time (but rather latencies in text), but also, since the `Attribute Assignment` is defined as "directly relating the respective pair of items or concepts" (ICOM/CRM Special Interest Group. 2020, 10)—with a built-in shortcut, so to speak, while we insist on detours. "Per default", the CRM states, all instances of properties described in a knowledge base "are the opinion of the team maintaining the knowledge base. This fact must not individually be registered for all instances of properties provided by the maintaining team, because it would result in an endless recursion of whose opinion was the description of an opinion." (ibid., 12)—However, we want to individually register these facts, and this recursion is kind of what we are after. Despite the great opportunities provided by CIDOC-CRM, and the prominence of the model, one has to give thought to a concern expressed by Jörg Wettlaufer: "Ontologies (just like controlled vocabularies) can't completely defuse the problem of subjectivity of world view … can we really grasp the humanities in a concept which was originally meant for a museum context?" (Wettlaufer 2018)

Consequently, a class is required that provides textual features, as well as subclasses for specification. The idea of building a conceptual hierarchy that covers every possible textual feature relevant in literary studies is one that could easily cause vertigo ('things to talk about when talking about a text' is more an abyss than a class). However, first efforts have been made: a superclass called `Receptional Entities`, itself a subclass of `Conceptual Object` with the subclasses `Semantic`, `Rhetoric` and `Formal Feature`. Subclasses to these three are only a provisional part of the ontology, and currently, externalising this into, for example, SKOS-classifications is being contemplated.

The last question in this regard is this: where are all these features actualized? This sometimes takes place in the text (like genre-specific features, or characters relevant throughout the text), most of the time in parts of the text—text passages. Consequently, a class for the text passage was introduced to the model (see Figure 4). On the other end of the spectrum, one might want to relate a text (or passage) not to another text, but to an abstract entity that is considerably larger than a text—an oeuvre perhaps, a discourse, an epoch. For these entities, which can all be seen as (or reduced to) text corpora, the class `Architextual Entity` was introduced (utilizing the term 'architext' as coined in Genette's fundamental study on intertextuality—Genette 1997), again as a subclass of `Conceptual Object`.

3.3 Universality

As mentioned above, the attribution of a feature, the identification of a relation can be modelled as a conceptual object. This is the case in the ontology proposed, but not in the important standard it is aligned to, CIDOC-CRM. At the heart of both CIDOC and FRBRoo, there are events. Both can be seen as divided into a static and a dynamic half, where the events from the dynamic half serve as the 'semantic glue' for all the other things (such as people, places, archive materials). The declaration/classification of an intertextual relation or any kind of text feature in this vocabulary does not rely on events (as a consequence of which one could call it only half an alignment). The thought behind this is simple: the ontology is meant to represent statements typically made in texts like commentaries, forewords, research papers, etc. These texts contain declarations and classifications, not as acts, which would be events in time, but as textual phenomena that probably, but not unconditionally, are the conserved results of events (intellectual acts), and ultimately are relevant only in their textual form. The conceptual bonus of this modelling lies in its integrative potential: no matter what one believes about how an `interpretament` got *into* a text—via an authorial intention, a certain reading, a contextual structure—we can all agree that somewhere there is a text.

This is one example of a path a vocabulary designed to include as many theoretical frameworks as possible has to choose: reduction, or, to be more precise: reduction of

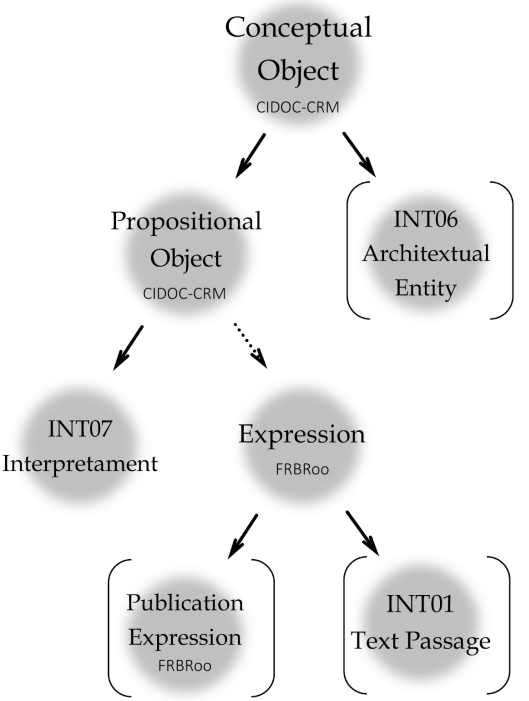

Figure 4: Notions of "Text" in hierarchical context.

elements that by default limit methodologies. As another example, one could point at the class `Type of Intertextual Relationship`. Instances of this class can have various methodological implications, and many of them could have been modelled more explicitly. However, this would have undermined the universal applicability intended by the ontology. Look at the difference between, say, Harold Bloom's theory of 'anxiety of influence' and the more recent theorem of the 'cultural archive' by Moritz Baßler. Both of these count as intertextuality theories (or theories including a subtheory concerned with intertextuality) but reside on quite distant points of the spectrum: whereas one of them captures a writer's resistance against his literary father-figures (Bloom 1973), the other one renders intertextual relations as, to put it in a nutshell, any kind of parallel between two texts that are part of the same

synchronous text corpus (Baßler 2005). Going for the exact representation of either of these two makes the representation of the other one difficult, if not impossible. While Bloom's theory, to name a significant difference, implies a notion of time and events, Baßler's does not. The proposed vocabulary has to take this wide variety of theories into account, by reducing all theoretical particulars under the notion of a simple 'Type of Intertextual Relationship,' which means: reducing them to instances of a class of this name (or, more likely: one of its subclasses or sub-subclasses).

3.4 Practicability: detail, granularity, and perspectivity

At this point, I'd like to return to the problem outlined at the beginning of this paper: the impossibility of distinguishing between commentary and interpretation. This is a fact that needs to be taken into account in any modelling of intertextual relations, all the more so for scholarly editions. The goal to produce a valid version of a text (whatever 'valid' means in the respective project context) goes hand in hand with the need to explicitly identify each act of interpretation as such—such as the declaration of an intertextual relation.

However, the following is also true: the fact that identifying intertexts is an interpretive act does by no means imply that an edition is obliged to do all the interpretation and model it in-depth. Providing knowledge on the edited text's intertexts is a service for a community of readers, who in turn bring their own interpretations to the text (see again Martens 1993, 49). Also, a practical matter must not be neglected: editions often deal with large quantities of texts, thus even larger quantities of intertextual relations (one may think of literary collages of quotations like T. S. Eliot's *The Waste Land* or Karl Kraus' *Third Walpurgis Night*), the in-depth analysis of which would exceed the scope (and, as one could argue, the purpose) of an edition project.

Every edition project concerned with intertextuality has to deal with this issue and find a solution from a wide spectrum of possibilities, ranging from the rudimentary to the in-depth modelling of intertextual relations. Thus, a vocabulary built for this purpose needs to provide flexibility.

As mentioned before, the modelling of intertextuality can take many forms, depending on what kinds of texts and what dimension of text (whole texts, text passages, text features, 'architexts') one wants to link and what additional information one wants to provide. This wide scope of different modelling options can be differentiated by three aspects: detail, granularity and perspectivity.

Perspectivity comes into play when a scholar's research, the edition or commentary itself, is included in the modelling. In that case, modelling the interpretive acts as `Interpretaments` (which then, in another step, could be linked to the scholarly discourse) would be useful (see Figure 5).

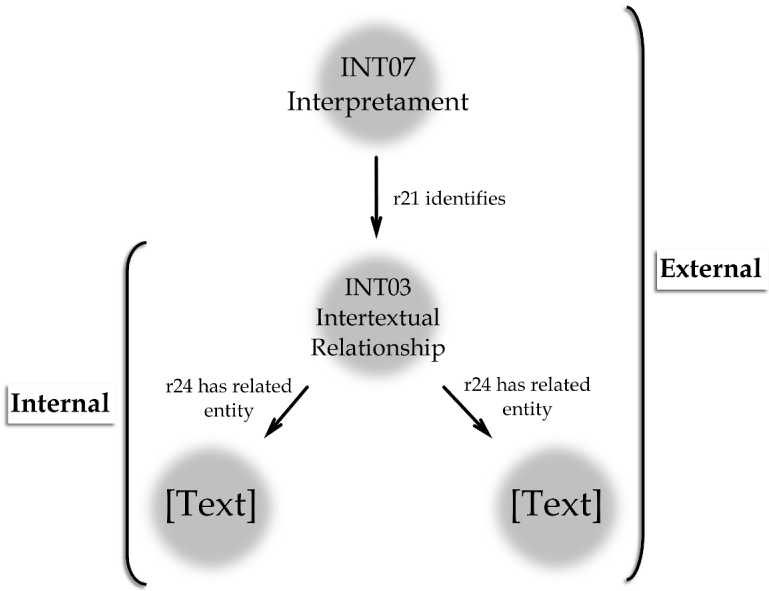

Figure 5: Internal and external perspectives on a scholarly act.

In terms of granularity, a single intertextual relation can be determined between two texts as a whole (low granularity) or numerous relations between their most minute fragments or features (high granularity). Finally, one can ask how much detail the projected model of, say, a digital edition's intertextual aspect should provide. The detail could range from the simple statement that an edited text refers to some source, to a modelling that includes a classification of the relation, to one that includes a source for a definition of said classification. Furthermore, it could be more specific in terms of what the intertextual relation actually relates to: the edited text as a whole, a text passage, or a certain feature shown by a passage (see Figure 6).

4 Conclusion

At more than one point in reading this paper, one may have been taken by surprise by yet another dissection of yet another aspect of the seemingly straightforward act of determining and classifying intertextual relations. (This experience then would mirror quite well the process of creating the vocabulary.) As the last remarks probably

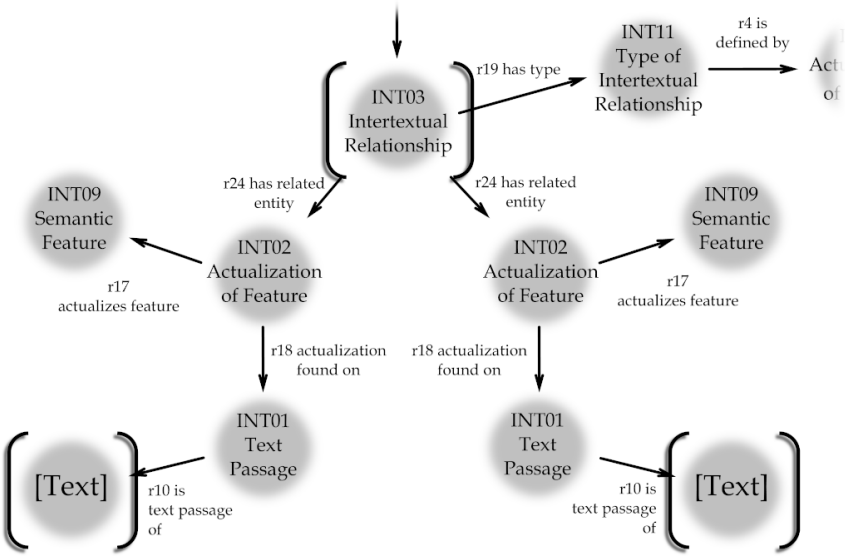

Figure 6: A modelling in higher detail; a lower detail modelling would make do with the three instances in brackets.

made clear, the vocabulary intends to provide the means for two things: to represent the respective findings and scholarly activities relatively in-depth—considering the sometimes hidden complexity of, for example, a literary scholar's day-to-day business—, but also, if necessary, to model intertextuality in a minimalistic way. The latter then could serve as a starting point for a gradual enrichment.

The proposed ontology *INTRO — an Intertextual Relationships Ontology for literary studies* (Oberreither 2018–) is in an advanced beta-state right now (early 2020). A current version is available online, as well as a test case, modelling a paragraph of one of Elisabeth Bronfen's 'cross mappings' and a documentation of central classes.

Bibliography

Bartalesi, Valentina and Carlo Meghini. 2017. "Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study." *Semantic Web Journal* 8 (3): 385-94. Accessed February 15, 2021. https://content.iospress.com/articles/semantic-web/sw198.

Bartlett, Sarah, and Bill Hughes. 2011. "Intertextuality and the Semantic Web: Jane Eyre as a test case for modelling literary relationships with Linked Data." *Serials: The Journal for the Serials Community* 24 (2): 160-5. doi:10.1629/24160.

Baßler, Moritz. 2005. *Die kulturpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie.* Tübingen: Francke.

—, and Rainer Karczewski. 2009. "Computergestützte Literaturwissenschaft als Kulturwissenschaft. Eine Wunschliste." *Jahrbuch für Computerphilologie* 9: 27–34. Accessed February 15, 2021. http://computerphilologie.digital-humanities.de/jg07/BassKarcz.html.

Bekiari, Chryssoula, Martin Doerr, Patrick Le Boeuf, and Pat Riva, eds. 2016. Definition of FRBRoo. A Conceptual Model for Bibliographic Information in Object-Oriented Formalism. Version 2.4. Accessed February 15, 2021. http://www.cidoc-crm.org/frbroo/sites/default/files/FRBRoo_V2.4.pdf.

Bloom, Harold. 1973. *The Anxiety of Influence. A Theory of Poetry.* New York: Oxford University Press.

Bronfen, Elisabeth. 2002. "Cross-Mapping. Kulturwissenschaft als Kartographie von erzählender und visueller Sprache." *Kulturwissenschaften. Forschung — Praxis — Positionen*, ed. by Lutz Musner, 110–34. Wien: WUV.

Ciotti, Fabio. 2014. "Digital literary and cultural studies: the state of the art and perspectives." *Between* IV (8). Accessed February 15, 2021. http://ojs.unica.it/index.php/between/article/view/1392/1183.

—. 2015. "Digital methods for Literary Criticism." [Presentation at the "Digital Humanities Autumn School 2015", Trier]. Accessed February 15, 2021. http://didattica.uniroma2.it/files/scarica/insegnamento/161783-Informatica-Umanistica-Lm-Per-Il-Llea/37175-Slide.

Connell, Sarah, and Julia Flanders. 2017. "Intertextual Networks: Theorizing and Encoding Textual Connections in Early Women's Writing." *Women Writers Project. The Blog*. Accessed February 15, 2021. http://web.archive.org/web/20210128174049/https://wwp.northeastern.edu/blog/dh2017.

Fuhrmann, Manfred. 1985. "Kommentierte Klassiker?" In *Warum Klassiker? Ein Almanach zur Eröffnung der Bibliothek deutscher Klassiker*, ed. by Gottfried Honnefelder, 37–57. Frankfurt am Main: Deutscher Klassiker-Verlag.

Genette, Gérard. Palimpsests. 1997 [1982]. *Palimpsests. Literature in the Second Degree*, trans. by Channa Newman and Claude Doubinsky. Lincoln: University of Nebraska Press.

Groddeck, Wolfram. 1993. "'Und das Wort hab' ich vergessen.' Intertextualität als Herausforderung und Grenzbestimmung philologischen Kommentierens, dargestellt an einem Gedicht von Heinrich Heine." In *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition, 4.–7. März 1992, autor- und problembezogene Referate*, ed. by Gunter Martens, 1–10. Tübingen: de Gruyter.

ICOM/CRM Special Interest Group. 2020. "Volume A: Definition of the CIDOC Conceptual Reference Model." Version 7.0.1. Accessed February 15, 2021. http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.0.1_%2018-10-2020.pdf.

Kanzog, Klaus. 1999. "Baustein, Kontext, Intertextualität. Zur Beziehung von Eigen- und Fremdtext als typologisches und editorisches Problem." In *Produktion und Kontext. Beiträge der internationalen Fachtagung der Arbeitsgemeinschaft für Germanistische Edition im Constantijn Huygens Instituut, Den Haag, 4. bis 7. März 1998*, ed. by Hendricus Theodorus Maria van Vliet, 329–35. Tübingen: Niemeyer.

Kristeva, Julia. 2003 [1967]. "Bachtin, das Wort, der Dialog und der Roman." In *Texte zur Literaturtheorie der Gegenwart*, ed. and commented by Dorothee Kimmich, Rolf Günter Renner, and Bernd Stiegler, 334–48. Stuttgart: Reclam.

Le Boeuf, Patrick, Martin Doerr, Christian Emil Ore, and Stephen Stead, eds. 2019. *Definition of the CIDOC Conceptual Reference Model. Volume A. Version 6.2.6.* Accessed February 15, 2021. http://www.cidoc-crm.org/Version/version-6.2.6.

Martens, Gunter. 1993. "Kommentar — Hilfestellung oder Bevormundung des Lesers?" *editio* 7: 36-50.

Oberreither, Bernhard. 2018– *INTRO — an Intertextual Relationships Ontology for literary studies.* Accessed February 15, 2021. https://github.com/BOberreither/INTRO.

—. 2019. "Zwei Überlegungen zur Konzeption einer Linked-Data-Ontologie für die Literaturwissenschaften." In *digital humanities austria 2018. empowering researchers*, ed. by Marlene Ernst et al., 134–9. 2020. doi:10.1553/dha-proceedings2018s134.

Peroni, Silvio, and David Shotton. 2018. "The SPAR Ontologies." *Submitted to the 17th International Semantic Web Conference. Proceedings, Part II.* Accessed February 15, 2021. https://w3id.org/spar/article/spar-iswc2018.

Pfister, Manfred. 1985. "Konzepte der Intertextualität." In *Intertextualität. Formen, Funktionen, anglistische Fallstudien*, ed. by Ulrich Broich and Manfred Pfister, 1–30. Tübingen: Niemeyer.

Pohl, Adrian. 2010. "Partizipativer Katalog, Intertextualität und Linked Data." *Übertext: Blog*. Last modified February 2, 2010. Accessed February 15, 2021. http://www.uebertext.org/2010/02/.

Rafferty, Pauline. 2016. "Genette, Intertextuality, and Knowledge Organization". Accessed February 15, 2021. https://www.researchgate.net/publication/290310957_Genette_intertextuality_and_knowledge_organization.

Sahle, Patrick. 2007. "Digitales Archiv — Digitale Edition. Anmerkungen zur Begriffsklärung." In *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, ed. by Michael Stolz, Lucas Marco Gisi, and Jan Loop. Accessed February 15, 2021. https://www.germanistik.ch/scripts/download.php?id=Digitales_Archiv_und_digitale_Edition.

—. 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik.* Norderstedt: Books on Demand.

Shotton, David. 2013. "Extending CiTO to enable use of the Open Annotation Data Model to describe citations." *Semantic Publishing.* Accessed May 5, 2021. https://semanticpublishing.wordpress.com/2013/07/03/extending-cito-for-open-annotations.

Stead, Stephen, Martin Doerr, Christian-Emil Ore et al. 2018. "CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation. Version 0.9 (Draft)." Accessed February 15, 2021. https://cidoc-crm.org/crminf/ModelVersion/version-0.9.

Stüben, Jens. 1933. "Interpretation statt Kommentar. Ferdinand von Saar und Marie von Ebner-Eschenbach: Kritische Texte und Deutungen." *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition, 4.–7. März 1992, autor- und problembezogene Referate*, ed. by Gunter Martens, 99–107. Berlin: de Gruyter.

TEI Consortium. 2020. "TEI P5: Guidelines for Electronic Text Encoding and Interchange." Accessed February 15, 2021. https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf.

Tennis, Joseph F. 2004. "URIs and Intertextuality: Incumbent Philosophical Commitments in the Development of the Semantic Web." In *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference*, ed. by Ia C. McIlwaine, 103–8. Würzburg: Ergon.

Van Zundert, Yoris. 2016. "Barely beyond the book." In *Digital Scholarly Editing. Theories and Practices*, ed. by Matthew James Driscoll, and Elena Pierazzo, 83–106. Cambridge, UK: Open Book Publishers.

Wettlaufer, Jörg. 2018. "Der nächste Schritt? Semantic Web und digitale Editionen." In *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, ed. by Roland S. Kamzelak and Timo Steyer. doi:10.17175/sb002_007.

Windfuhr, Manfred. 1991. "Zum Verständnis von Kommentar und Genese." *editio* 5: 173–7.

Woesler, Winfried. 1997. "Der Autor und seine Quellen aus der Sicht des neugermanistischen Editors." In: *Quelle — Text — Edition. Ergebnisse der österreichisch-deutschen Fachtagung der Arbeitsgemeinschaft für germanistische Edition in Graz vom 28.2.–3.3.1996*, ed. by Anton Schwob, Erwin Streitfeld, and Karin Kranich-Hofbauer, 3–19. Berlin: de Gruyter.

Zittel, Claus. 1997. "Von den Dichtern. Quellenforschung versus Intertextualitätskonzepte, dargestellt anhand eines Kapitels aus Friedrich Nietzsches *Also sprach Zarathustra*." In *Quelle — Text — Edition. Ergebnisse der österreichisch-deutschen Fachtagung der Arbeitsgemeinschaft für germanistische Edition in Graz vom 28.2.–3.3.1996*, ed. by Anton Schwob, Erwin Streitfeld, and Karin Kranich-Hofbauer, 315–31. Berlin: de Gruyter.