Georg Rehm

Observations on Annotations

Abstract: The annotation of textual information is a fundamental activity in Linguistics and Computational Linguistics. This article presents various observations on annotations. It approaches the topic from several angles, including Hypertext, Computational Linguistics and Language Technology, Artificial Intelligence and Open Science. Annotations can be examined along different dimensions. In terms of complexity, they can range from trivial to highly sophisticated, in terms of maturity from experimental to standardised. Annotations can be annotated themselves using more abstract annotations. Primary research data such as, e.g., text documents can be annotated on different layers concurrently, which are independent but can be exploited using multi-layer querying. Standards guarantee the interoperability and reusability of data sets. The chapter concludes with four final observations, formulated as research questions or rather provocative remarks on the current state of annotation research.

Keywords: Evaluation, Levels of Annotation, Markup, Semantic Web, Artificial Intelligence, Computational Linguistics, Digital Humanities, Digital Publishing

1 Introduction

The annotation of textual information is one of the most fundamental activities in Linguistics and Computational Linguistics including neighbouring fields such as, among others, Literary Studies, Library Science and Digital Humanities (Ide and Pustejovsky 2017; Bludau et al. 2020). Horizontally, data annotation plays an increasingly important role in Open Science, in the development of NLP/NLU prototypes (Natural Language Processing/Understanding), more application- and solution-oriented Language Technologies (LT) and systems based on neural technologies in the area of Artificial Intelligence (AI).

This article reflects on more than two decades of research in the wider area of annotation including multi-layer annotations (Witt et al. 2007a,b), the modelling of linguistic data structures (Wörner et al. 2006; Rehm et al. 2007b; Witt et al. 2009) including hypertext and web genres (Rehm 2002, 2007, 2010a), the production and distribution of annotated corpora (Piperidis et al. 2014; Rehm 2016; Rehm et al. 2020a) and the use of metadata, annotation schemes and markup languages

Georg Rehm, DFKI GmbH, georg.rehm@dfki.de

https://doi.org/10.1515/9783110689112-014

(Rehm et al. 2008a,b, 2009; Rehm 2010b). After an initial approximation of a definition (Section 2), the chapter provides lessons learned, future research directions as well as observations on the scientific and technical process of annotating textual data from several angles including Hypertext, Markup and the World Wide Web (Section 3), Computational Linguistics (Section 4), Artificial Intelligence (Section 5), Language Technology (Section 6) and Open Science (Section 7). The article concludes with an overview of the main conceptual dimensions involved in the annotation of textual information (Section 8) and a summary (Section 9).

2 Definition

Definitions of the term "annotation" typically focus on either procedural (i.e., process-related), technical (i.e., markup-related) or conceptual (i.e., semantics-related) aspects, sometimes also combinations of the different layers (Goecke et al. 2010; Ide and Pustejovsky 2017). The notion we follow in this article is loosely based on the concept of Annotation Graphs (Bird and Liberman 2001), which can be used to represent an unlimited number of annotation layers, while links between the text and annotations can be established in an unrestricted way (Witt et al. 2007b; Ide and Suderman 2007). Specifically, we view annotations as *secondary research data* added to *primary research data*. Annotations are, therefore, part of the metadata that also include general information on the primary data (author/creator, modality, creation date etc.). *Linguistic* annotations, then, cover "any descriptive or analytic notations applied to raw language data. The basic data may be in the form of [...] audio, video and/or physiological recordings [...] or it may be textual. The added notations may include transcriptions of all sorts (from phonetic features to discourse structures), part-of-speech and sense tagging, syntactic analysis, 'named entity' identification, co-reference annotation, and so on." (Bird and Liberman 2001). The procedure of annotating data can include, among several other variants, highlighting and labelling specific segments, commenting upon certain aspects, and selecting as well as inserting markup elements (tags) into a text document. The design of a concrete annotation scheme typically follows at least two consecutive phases: based on linguistic theory or insights, an annotation model is created (Pustejovsky et al. 2017) for which, then, a technical representation is developed (Ide et al. 2017b). Finlayson and Erjavec (2017) provide an overview of the processes and tools involved in the creation of annotations.

3 Hypertext, Markup and the World Wide Web

Annotations have always been an integral concept of hypertext (Nelson 1987) itself as well as the World Wide Web. In his seminal piece, "As we may think", Bush (1945) described his vision of the Memex, explaining that the user of the Memory Extender "can add marginal notes and comments [...] by a stylus scheme". And Berners-Lee (1989) described, in the original concept note that laid the groundwork for what later became the World Wide Web, that one "must be able to add one's own private links to and from public information. One must also be able to annotate links, as well as nodes, privately." While Berners-Lee had this specific idea in mind already back in 1989, it took more than 20 years of work for Web Annotations to become a web standard proper (see below).

Linguistic annotations are, procedurally, conceptually, and technically, closely linked to markup and markup languages, especially the ones based on XML (Extensible Markup Language, Bray et al. 2008), enriched, processed, presented and queried with related formalisms such as, among others, XML Schema, XSLT, XPath, XQuery, CSS, RDF and OWL. Through their unambiguous, syntactic separation of annotations from the primary data, markup languages are a natural candidate for linguistic annotations, especially those based on XML, the most widely used metalanguage for the definition of concrete markup languages using approaches such as XML Schema or Document Type Definitions (DTD). One of the most widely used annotation systems in Linguistics and Digital Humanities are the TEI guidelines (TEI Consortium 2019), initially developed in the late 1980s. The formalisms mentioned above were developed and standardised by the World Wide Web Consortium (W3C), an international non-profit organisation founded by Tim Berners-Lee in 1994 to lead the further development of the World Wide Web's technical building blocks. Just like XML, the W3C's effort to move from a static, document-centric to a *Semantic* Web also lead to a number of highly influential and innovative developments in Linguistics and Computational Linguistics, especially with regard to modelling and querying annotations (Rehm et al. 2007a; Farrar and Langendoen 2010; Chiarcos and Sukhareva 2015). The interface between technical markup and linguistic annotations is examined by Metzing and Witt (2010) including the interface between HTML and linguistic markup (Rehm 2010a).

Most stand-alone tools for the annotation of linguistic data, often implemented in Java, have by now vanished or, if they are still in use, target a specific niche for which a browser-based solution has not been developed yet. Nowadays, actual annotation work is typically carried out in the web environment, i.e., in the browser, using one of the web-based annotation tools such as, among others, Brat (Stenetorp et al. 2012), WebAnno (Eckart de Castilho et al. 2016), INCEpTION (Klie et al. 2018)

or CATMA (Meister et al. 2019). Crucially, the textual data that is annotated this way may be web data (i.e., HTML documents) that was downloaded or crawled, but it is typically *not* live web data because anchoring annotations to live web documents that can change, in a subtle or substantial way, any minute is technically challenging.

The fairly recent W3C standard Web Annotation was developed for exactly this purpose, i.e., to enable the annotation of live web data. The standard consists of three W3C recommendations. The Web Annotation Data Model (Sanderson et al. 2017b) describes the underlying annotation data model as well as a JSON-LD serialisation. The Web Annotation Vocabulary (Sanderson et al. 2017c) underpins the Data Model, and the Web Annotation Protocol (Sanderson 2017a) defines an HTTP API for publishing, syndicating and distributing Web Annotations. The standard enables users to annotate arbitrary pieces of web content in the browser, essentially creating an additional, independent layer on top of the regular World Wide Web. Web Annotations are *the* natural mechanism to enable web users and readers, on a general level, interactively to work with content, to include notes, feedback and assessments, to ask the author or their peers for references or to provide criticism. However, there are still limitations. As of now, none of the larger browsers implement Web Annotations natively, i.e., content providers need to enable Web Annotations by integrating a corresponding JavaScript library. Another barrier for the widespread adoption of Web Annotations are proprietary commenting systems, as used, among others, by all major social networks who are keen on keeping all annotations (i.e., comments and other types of user-generated content) in their own respective silos and, thus, under their own control.

Nevertheless, services such as the popular Hypothes.is tool (see below) enable Web Annotations on any web page, but native browser support, ideally across all platforms, is still lacking. In addition to the (still somewhat limited) ability of handling live web data, the Web Annotation standard has multiple advantages that make it perfectly suited for linguistic annotations. The Web Annotation Data Model is very general and can be conceptualised as a multi-layer Annotation Graph. Annotations are sets of connected resources, typically an annotation *body* and the *target* of the annotation. If and when the Web Annotation standard is finally available natively in all browsers, conversations between users and content creators can take place anywhere on the web in a standards-compliant way, where, and this is crucial, the annotations are under the control of the users because annotations can live separately from the documents they are pointing to – they are reunited and re-anchored in real time.

The annotation tool developed by the non-profit organisation Hypothes.is is by the far the most popular one. It enables taking private notes or publishing public annotations. It can be used in collaborative groups, it provides Linked Data

connections and works with different formats including HTML, PDF and EPUB. It is used in scholarly publishing and as a technical tool for open peer review, in research, education and investigative journalism.[1] It can also be used for automated annotations, e.g., to tag Research Resource Identifiers (RRIDs).

With regard to the current state and further development of markup approaches and technologies, XML, originally published in 1998 and, since then, in extremely widespread use, is no longer actively maintained or developed further within W3C. However, there is still a highly active and passionate community interested especially in declarative markup. Discussing some of the lessons learned during the development of XML, Walsh and Bethan (2018) emphasise the need for a new umbrella environment and community initiative for future work on descriptive markup: the Markup Declaration.[2]

4 Computational Linguistics

The annotation landscape, which consists, generally speaking, of tools and formats, has had several decades to grow and to mature into an area that is impossible to characterise in the context of a short book chapter alone. Many colleagues provided general or specific overviews, including, among others, Bird and Liberman (2001), Dipper et al. (2004), Metzing and Witt (2010), Stührenberg (2012), Ide and Pustejovsky (2017), Biemann et al. (2017), Stede (2018), Neves et al. (2019). In addition to a large number of all-purpose and specialised formats (Ide et al. 2017a) such as, among many others, TEI, NIF, NAF, LAF, GRAF, TIGER, STTS, FoLIA, there is a plethora of editors and tools to chose from, such as Brat, WebAnno, Exmaralda, Praat, ELAN, ANNIS, CATMA, INCEpTION and Prodigy as well as many others including crowd-sourced approaches.

Both annotation tools and also annotation formats can be described along a number of dimensions and continuums. Annotation schemes range from *trivial* (e.g., marking up single tokens) to *complex* (enabling semantically deep and nuanced annotations). These often correlate with their annotation task, from *easy*, *straightforward* and *well understood* (e.g., annotating named entities) to *hard*, *challenging* and *novel* (e.g., the annotation of actors and events in storylines). Accordingly, simple annotation tasks, the goals of which can be summarised and specified in concise annotation guidelines effectively, typically result in very high

1 See, for example, the projects presented in the various events of the "I Annotate" conference series, which started in 2013: http://iannotate.org (04.02.2020).

2 https://markupdeclaration.org (04.02.2020)

inter-annotator agreement scores while hard, ambitious and challenging tasks that may require a certain level of expertise or training, rather result in low inter-annotator agreement (Gut and Bayerl 2004; Bayerl and Paul 2007, 2011; Snow et al. 2008; Artstein 2017). Finally, simple annotation tasks are typically carried out using general all-purpose tools while complex annotation tasks usually require specialised or customised tools.

5 Artificial Intelligence

Artificial Intelligence (AI) as an academic discipline was founded in the 1950s. While it consists of various subfields, by now, it is ubiquituous first and foremost due to the recent breakthroughs made in the area of Machine Learning (ML) using Deep Neural Networks (DNNs). These have been made possible due to powerful supervised but also unsupervised machine learning algorithms, fast hardware and, crucially, large amounts of data. This is why the relevance of annotations and annotated data sets for AI at large, including Language-Centric AI (Rehm et al. 2020d), i.e., Computational Linguistics and Natural Language Understanding, has increased dramatically in recent years.

Modern AI methods are data-driven. Supervised learning methods rely on very large annotated data sets, many of which consist of primary (language) data and secondary annotations, as defined in Section 2.[3] In fact, data curation and annotation has become so important that new business models have emerged that revolve around the production of structured data for customers who want to make use of supervised learning in concrete application scenarios. Some companies employ in-house experts for the construction of data sets while others use crowd-working approaches.[4] Key aspects of any data generation process include the annotation speed, the quality and relevance of the annotations, and how meaningful, reliable and representative the annotations are.

With regard to the context of AI-based applications, the line between the construction of structured data sets on the one hand and the collection of – typically user-generated – data points on the other, is blurry, as both can be conceptualised as annotations. In the former, language data is annotated with regard to, for ex-

3 In Natural Language Understanding, DNNs are also used for language modelling, i.e., for generating statistical models out of enormous amounts of unannotated language data. These can be used for various classification and prediction tasks (Ostendorff et al. 2019).

4 For example, Appen's current slogan is "Data with a human touch: High-quality data for machine learning, enhanced by human interaction" (https://appen.com [04.02.2020]).

ample, word senses or intents. In the latter, actual live content is "annotated", for example, by liking a tweet, leaving a five-star rating for a restaurant or commenting on a news article. All of these activities are annotations that add metadata to existing data. Clicking a headline to go to an article or even turning the page in an ebook can also be and, in fact, are interpreted as annotations with regard to the underlying primary data in question. Increasingly slower page turns in an ebook, for example, could be interpreted by the user modelling algorithm as "boredom" with the current chapter and may, later on, result in automatically adjusted book recommendations. Even the non-action of no longer reading an ebook can be seen as an "implicit" annotation. In the future, for certain non-fiction genres it will be possible to identify the chapters in which readers lose interest and then to generate slightly different versions or paraphrases of those chapters with the intent of not losing any readers by keeping their engagement high. In these cases, the original human author will compete with the machine in an A/B test, i.e., both variants are presented to users in a short experimental phase, while only the statistically more effective variant will be used in the long-term. In today's digital age, users of large online applications must be aware of the fact that every single action or click they perform, i.e., every single annotation, is recorded, associated with their profile, and made use of by user modelling and recommender algorithms, including advertisements.

6 Language Technology

The applied field of Language Technology (LT) transfers theoretical results from language-oriented research into technologies and applications that are ready for production use. Linguistics, Computational Linguistics, Psycholinguistics, Computer Science, AI and Cognitive Science are among the relevant fields made use of in LT-solutions. Spell checkers, dictation systems, translation software, search engines, report generators, expert systems, text summarisation tools and conversational agents are typical LT-applications.

This Section takes a brief look at potential ways how LT as well as AI can interface with the Web Annotation technology stack (Section 3). LT can be embedded in various phases and places of the Web Annotation workflow to address and eventually solve a number of common challenges (Rehm et al. 2016). First, the web content to be enriched with annotations can be created automatically or semi-automatically using Natural Language Generation (NLG) approaches; in fact, this is already the case for vast amounts of online content, including online shops, weather reports, and articles about sport events. Second, the web content can be

automatically analysed and then annotated using LT, for example, for the purpose of generating an abstract of a longer article using automated text summarisation and then presenting the article to users in the form of an annotation. Third, the content of the actual annotations, potentially made by many different users, can be analysed using LT, for example, for the purpose of mining the feedback of the users or readers for sentiments and opinions towards the primary content, which may be a product description, a news article on a breaking event or a discussion of a topic of high social relevance. In that regard, web annotations are also – just like blogs, online videos, online photos – User-Generated Content (UGC). Currently, with individual silos containing UGC, it is complex, challenging and costly to perform Social Media Analytics and Opinion Mining at scale due to the various formats and heterogeneous sources. A centralised approach based on Web Annotation would simplify such text mining approaches significantly, also enabling a much broader and more varied analysis of opinions regarding, among others, commercial products, societal challenges, political trends and misinformation campaigns (Moreno-Schneider et al. 2017; Rehm 2018; Rehm et al. 2018a,b).

The Web Annotation standard is based on the notion of stand-off annotation, i.e., the annotations are not embedded inline within the actual primary data in the form of, e.g., XML elements, but stored indepedently from the primary data. This approach enables overlapping annotations, i.e., stand-off annotations do not have to adhere to the rather strict requirements regarding the tree structure imposed by the XML standard. Instead, stand-off annotations make use of a pointing or linking mechanism so that an annotation is anchored to or linked to a certain sequence of primary data. This (important) advantage comes with a computational cost, though, because each stand-off annotation needs to be explicitly anchored at processing time. In our recent and current research projects[5] we use a similar approach, the NLP Interchange Format (NIF, see Hellmann et al. 2013). NIF was developed especially for LT applications and is based on the Linked Data paradigm, i.e., RDF and OWL.

Between the development phase and the deployment phase of an LT-based solution, annotation formats can also be mixed. For example, in LYNX, all processing solutions make use of NIF (Rehm et al. 2019) but during the development and training phase of the German Legal NER model we used the CONLL format which is a simple, tab-seperated value, i.e., non-XML-based inline annotation format (Leitner et al. 2019, 2020).

5 DKT (http://digitale-kuratierung.de [04.02.2020]) (Bourgonje et al. 2016), QURATOR (https://qurator.ai [04.02.2020]) (Rehm et al. 2020b) and LYNX (http://lynx-project.eu [04.02.2020]) (Rehm et al. 2019).

7 Open Science

The umbrella term Open Science denotes the movement to make scientific research, data and dissemination accessible to interested stakeholders. It includes a multitude of different aspects, e.g., publishing open research, pushing for Open Access (instead of closed) and encouraging researchers of all fields to publish not only their results but also their data for easier verification and reproducibility. Open Science is becoming more and more popular and is, crucially, relevant to the broader topic of annotations. If we examine the taxonomy[6] produced by the EU project FOSTER to describe the different aspects of Open Science, these connections become immediately apparent: Open Science advocates for Open Data, which should not only be open but also annotated using standards, made available using platforms that are accessible (e.g., Linked Data) and described with metadata and semantics including well defined categories and taxonomies.

One of the key goals of promoting Open Research Data is to enable data re-use and, thus, Open Reproducible Research that also includes Open Science Workflows, often made possible by distributing Open Source software and specifying the workflows used to arrive at the results published in a scientific article. Annotations, the meaning and semantics of which are clearly documented, ideally using international standards, are the glue between the software components that produce the annotations, annotated open research data, annotation guidelines, research data repositories, query mechanisms and scientific publications.

With the ever growing and maturing technology infrastructure for data-intensive research, Open Science will soon become the norm, including the use of sustainable repositories for making available research data clearly described and annotated using standardised, best-practice approaches, linked to other sets of research data, fostering the re-use of the data in the context of new research questions. The FAIR Data Principles emphasise, in their procedural order, four main aspects of research data, which should be made findable, accessible, interoperable and re-usable (Wilkinson et al. 2016).[7] Most of the FAIR principles refer to metadata, which can, especially if they relate to primary data, also be conceptualised as annotations. The relevant principles are the following ones:

F2 Data are described with rich metadata.
F3 Metadata clearly and explicitly include the identifier of the data they describe.

6 See https://www.fosteropenscience.eu/foster-taxonomy/open-science (04.02.2020).
7 See https://www.go-fair.org (04.02.2020) for more detailed descriptions of the principles.

A1 (Meta)data are retrievable by their identifier using a standardised communications protocol.
A2 Metadata are accessible, even when the data are no longer available.
I1 (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2 (Meta)data use vocabularies that follow FAIR principles.
I3 (Meta)data include qualified references to other (meta)data.
R1 (Meta)data are richly described with a plurality of accurate and relevant attributes.
R1.2 (Meta)data are associated with detailed provenance.
R1.3 (Meta)data meet domain-relevant community standards.

As can be seen, the FAIR principles – and also Open Science in general – recommend, at their core, the use of standards for the purpose of enabling or enhancing, as much as possible, the findability, accessibility, interoperability and reusability of research data (see Labropoulou et al. 2020, for a practical example). While these recommendations are important and, thus, to be supported, it is also worth noting that especially basic research is about trying and inventing *new* things, i.e., things that have, almost by definition, *not* been standardised yet. This contradicts, on a fundamental level, with the recommendation of using standards as the consensus reached within a specific research community to represent, for example, temporal expressions in natural language text. The contradiction can be resolved, though, if the recommendation is relaxed to the use of established tools and best practice approaches as well as the modification and extension of standards. The crucial aspect is to document the semantics of the annotation scheme used in a corpus or data set. If an established, standardised approach does not work for an emerging piece of research, a new approach needs to be created or an established approach modified.

It is safe to predict that Open Science will be transforming research in the next years, making it more sustainable, more visible and more transparent. Several disciplines have already been following Open Science-like approaches for quite a while. On a larger scale, though, Open Science will only be fully possible with substantially improved digital infrastructures. Notable initiatives are the European Open Science Cloud (EOSC)[8] and the Nationale Forschungsdateninfrastruktur (NFDI)[9] in Germany. Additionally, we can predict that, soon, robust and large-scale services for the annotation of documents will be provided, starting with scien-

8 https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud (04.02.2020)
9 https://www.dfg.de/foerderung/programme/nfdi/ (04.02.2020)

tific publications, for which it will be possible to annotate and, thus, explicitly represent, using standardised metadata schemas and ontologies, their methods used or expanded upon, evaluation approaches, data sets as well as findings and contributions – this structured set of semantic information associated with one research article, as the atomic unit of scientific publication, will be contextualised in larger knowledge graphs which will capture the research output of entire scientific fields, including annotations. Several larger scientific publishing houses are already now developing corresponding digital infrastructures to capture the results they publish. At the same time, the Open Research Knowledge Graph (ORKG) initiative promotes the vision of moving scholarly publishing from a coarse-grained, predominantly *document-based* to a *knowledge-based* approach by, first, automatically identifying and extracting and, second, representing and expressing scientific knowledge through semantically rich, interlinked graphs (Jaradeh et al. 2019).[10] In a third step, the knowledge contained in the ORKG can be used, for example, to compare the approaches followed in different scientific papers on the same research question.

8 Dimensions of Annotations

The process of adding annotations to a set of primary research data can be conceptualised as the insertion of secondary research data (see Section 2). The secondary data added to the primary data typically refers to one or more (often interconnected) properties of the primary data that are explicitly marked using syntactically identifiable methods. Figure 1 on the next page shows the general aspects and dimensions involved in an annotation in more detail; Ide and Romary (2001) provide a similar but more technical view focused upon syntactic annotations.

An annotation explicitly describes a *property* of a piece of primary data using a tuple that consists of the *label* of the property in question (e.g., "part of speech") and a corresponding *value* (e.g., "adjective"). An annotation can also include a pointer to an abstract, internally or externally represented annotation scheme that, typically, specifies the semantics of all possible annotations. This annotation scheme, in turn, can be used to constrain or to restrict specific annotations, i.e., the <label, value> pair that makes up an annotation.

Especially when designing a new or modifying an existing annotation scheme to address a specific research experiment, several relevant questions need to be taken into account, some of which are included in Figure 1 on the following page.

10 https://www.orkg.org (04.02.2020)

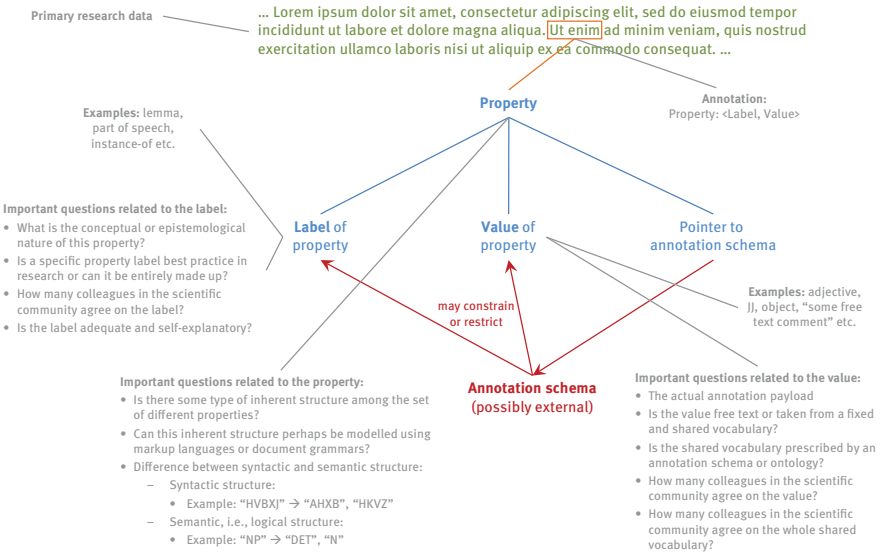

Fig. 1: General aspects and dimensions of annotations

These question pertain, among others, to the conceptual or epistemological nature of the specific label of an annotation: on the one hand, this label can denote a concept that has been established in a scientific field for decades or it can refer to a fairly recent aspect, phenomenon or finding, for which an established term in the respective scientific community does not exist yet. Another aspect relates to the set of properties that are being described with the help of an annotation scheme: are these isolated properties without any inherent structure that governs the sequence or distribution of their instantiations (e.g., different types of named entities) or does some kind of linguistic or syntactic structure exist on top of the different annotations? If the latter is the case, can this structure be explicitly modelled, for example, using mechanisms built into XML DTD or XML Schema-based document grammars (Maler and El Andaloussi 1996; Megginson 1998)? Can, maybe as an additional mechanism on top of the document grammar, an ontology be used to describe higher-level semantic concepts?

The various notions hinted at in Figure 1 lead us to a more abstract aspect of annotations: just like primary research data, annotations have various properties themselves. Depending on the research question and overall use case, it may be important or even necessary to explicitly represent these properties, i.e., to annotate annotations. Among this set of properties are the following: *annotator* of the annotation (i.e., was it created by a human expert or by an automatic process?), *annotation layer* (i.e., does the annotation refer to the "document struc-

ture", "layout", "syntax", "semantics", "information structure" etc.?), *confidence value* (i.e., how confident is the human annotator or automated process that an annotation is correct?), *timestamp* (i.e., when the annotation was added), *style* (i.e., how an annotation is rendered in a certain system) and *application scenario* (i.e., is the annotation primarily meant for human or machine consumption?). It is important to note that more structure can be explicitly added even on top of these annotations, especially with regard to the relationship and interdependence of the various annotation layers.

Instantiated sets of annotations can be described along various axes and dimensions, some of which are rather vague while others are more concrete.

- *Annotator:* The actual source or origin of annotations included in a data set, for example, one or more automated components, human experts, human laypersons, crowd workers etc. This dimension also refers to the *methodology* followed for including the annotations into the primary data.
- *Semantics:* The semantics of the annotations, i.e., the nature of the properties explicitly and formally described through the annotations, e.g., linguistic concepts or aspects relating to document structure, rhetorical structure, genre, style, terminology etc. This dimension is connected to the *annotation scheme* used, which could be an experimental scheme developed, e.g., in a research project for a novel purpose, or one of the well known annotation schemes and standards that have been in use for decades, e.g., TEI.
- *Layers:* The nature and interconnectedness of the different annotation layers if an annotated data set contains multiple layers.
- *Guidelines:* A crucial question with regard to annotation projects primarily carried out by humans, relates to the presence of annotation guidelines, especially with regard to the specification of concrete examples and exceptions, i.e., which concepts to annotate how in a specific context.
- *Research question or application use case:* An annotated data set is typically associated either with an underlying research question that has motivated the construction of a data set or with a concrete annotation pipeline (i.e., application use case) that was used to annotate the primary data.
- *Complexity:* This dimension refers to the notion that some annotations are more complex than others, it is closely related to several other dimensions.
- *Evaluation:* Most annotated data sets have been evaluated in some way, e.g., with regard to the inter-annotator agreement (if the primary data was annotated by multiple annotators).

Space restrictions prevent us from describing all dimensions in more detail, which is why we concentrate on *Complexity* (Section 8.1) and *Evaluation* (Section 8.2).

8.1 Complexity of Annotations

In Computational Linguistics and also in the wider Digital Humanities area, several fairly detailed annotation schemes and markup languages have been developed for the annotation of textual data in the last 30 years. The TEI guidelines are probably the most extensive ones – the PDF version of the TEI P5 guidelines (TEI Consortium 2019) has a length of almost 2000 pages, in which hundreds of XML elements and attributes, grouped into various modules, are described. In stark contrast, the annotation schemes used in many current data sets, especially for large-scale, data-driven AI approaches that rely on vast amounts of training data, are quite shallow and highly generalised. Machine learning approaches perform best with large amounts of training data; it is beneficial for the performance of the resulting models and classifiers if the number of unique class labels is rather small and the number of different examples per class label rather high. Especially for environments in which such AI-based classifiers are used in production, the corresponding data sets are often created by professional annotation teams or companies (see Section 5). In these scenarios and use cases it is not feasible to annotate data sets with complex annotation schemes.

It is an interesting question for future research if the difference in complexity or the "level of sophistication" of different annotation schemes – from a simple set of a few labels to highly complex markup languages like TEI P5 – can be measured or formally described. To the best of the author's knowledge, there has not been any work on this topic so far. Many different data points and statistics about an annotation scheme could be exploited for this purpose, e.g., the number of property labels (i.e., XML tags), the number of meta properties (e.g., XML attributes), the number of free text and predefined values, the presence of inherent structure including nesting levels etc. These, and other, statistics could be included in a formula that captures the complexity of an annotation scheme; it could also be used, together with data such as token/annotation ratio, to model the complexity of the annotations contained in a concrete data set.

8.2 Evaluation of Annotations

The evaluation of annotations is a crucial dimension of formally describing a data set or corpus, especially when it was created for the purpose of training a practical tool and also when an emerging annotation scheme was used. In that regard, two different aspects can be evaluated that are intricately interrelated: the annotation scheme itself and concrete annotations.

The evaluation of the validity of an abstract, possibly emerging, annotation scheme is typically an iterative process (Dickinson and Tufiş 2017; Artstein 2017): first, an initial version of the annotation scheme is applied to a small and, ideally, representative data set to examine if it is practical and balanced concerning its ability to annotate all the characteristics and phenomena it is supposed to be able to mark up explicitly. An overarching aspect that should be taken into account when developing and iteratively evaluating an annotation scheme relates to the question if it models scientific consensus. These initial tests are, later on, repeated with more mature versions of the annotation scheme until all requirements, prescribed by the respective research question, are met. As the two go hand in hand, these initial evaluations typically concern not only the annotation scheme but also the annotation guidelines as well as their applicability using a specific annotation tool. Important questions regarding the annotation guidelines relate to their length, coverage, examples, and exceptions as well as how long it usually takes to train annotators so that they can perform an annotation task.

The result of an annotation task or process can also be evaluated, both qualitatively and quantitatively. In the context of this chapter, the typical approach is to compare multiple annotations of the same primary data, created by multiple annotators, and to compare their inter-annotator agreement, i.e., how well do the various annotators agree when comparing their respective annotations. Multiple approaches to calculate inter-annotator agreement exist (Gut and Bayerl 2004; Bayerl and Paul 2007, 2011). This analysis is crucial for data and experiment-related aspects such as replicability and reproducibility and for measuring the consensus among the annotators, especially for complex annotation tasks or emerging annotation formats. A variation of measuring inter-annotator agreement can be described as "intra-annotator agreement", i.e., the same annotator is asked to perform the same annotation task multiple times but under different conditions or several days or weeks apart. This approach can also be used to identify weaknesses in emerging annotation schemes or guidelines.

9 Summary and Conclusions

This article presents various observations on annotations. It approaches the topic from multiple angles including Hypertext, Computational Linguistics and Language Technology, Artificial Intelligence and Open Science. Annotations can be examined along different dimensions. In terms of complexity, they can range from trivial to highly sophisticated, in terms of maturity from experimental to standardised. Annotations can be annotated themselves using more abstract annotations.

Primary research data such as, e.g., text documents can be annotated on different layers concurrently (e.g., general segmentation including text structure, coherence relations, syntax), which are independent but can be exploited using multi-layer querying. Standards guarantee interoperability and reusability of data sets, which is especially crucial in terms of Open Science.

The chapter concludes with four final observations, formulated as research questions or rather provocative remarks on the current state of the field.

Do standards hold back innovative annotation research? Standard annotation schemes represent the condensed consensus gathered within a wider research community regarding certain phenomena. This class of standardised formats is crucial for interoperability and reproducibility. However, one aspect that is often neglected concerns the fundamental nature of research itself, which is about finding, creating and inventing *new* things, new pieces of knowledge, new insights, including new ways of annotating language data. Especially taking into account those annotation schemes that are, both conceptually and also technically, highly similar, it is worth emphasising that new breakthroughs require new approaches. Focusing on standards too much may hold back research.

Can we concentrate on annotating live web data instead of dead web data? Primary research data is nowadays typically annotated within a web-based environment, i.e., using a dynamic web application that visualises both the primary and the secondary research data in a browser. Very often, said primary data is, in fact, web data, i.e., text or multimedia data that was either crawled or collected using other means from the World Wide Web. Crawling and archiving live web data decouples the documents from their natural habitat, which essentially results in frozen snapshots of these documents. While this approach has been best practice in Computational Linguistics almost since the beginning of the World Wide Web, it would be much more interesting to treat the *live* World Wide Web as a corpus. Given that the web technology stack even includes its own annotation approach (Web Annotation, see Section 3), we should attempt to treat the whole, live World Wide Web as a giant corpus by parsing the whole web and by adding linguistic information using the Web Annotation approach, which can then be queried for linguistic analyses or for training machine learning models (Rehm 2018; Rehm et al. 2018a). To that end, larger collections of web-native Language Technology services (Rehm et al. 2020a,b) could be used in high-performance infrastructures (Rehm et al. 2020c).

Is it possible to design a machine-readable packaging format for describing annotations? Annotations have different dimensions along which they can be described (Section 8). It would be a highly interesting question to examine if it is possible to design a compact, machine-readable packaging format for describing annotation projects including the annotations themselves as well as the overall

approach, main formal aspects of the annotation scheme (including its complexity) and the concrete annotations. This is a relevant and important question from the point of view of Open Science (and more transparent as well as reproducible and interoperable science). The question also relates to machine learning, language resources and emerging AI and LT platforms. Soon, these will be able to import a data set and use a machine learning toolkit automatically to train a new model (Rehm et al. 2020c). In order for this to work fully automatically, we need metadata schemes to describe annotated data sets including formal aspects such as their annotation schemes and involved dimensions.

Is the field ignoring decades of valuable annotation science research? Since the emergence of the first large corpora and the statistical turn in the early 1990s, Computational Linguistics has produced a plethora of results and insights regarding the annotation of language resources – so much so that Ide (2007) even speaks of "annotation science". In the last five years, neural approaches have turned out to be very popular in Language Technology, outperforming essentially all of the previous methods. Generally speaking, neural technologies require very large data sets for training models. Corresponding applications are often generalised as classification tasks that are based on large data sets that were annotated with only few labels. In many cases, both the classification tasks and also the sets of labels or annotations must be described as rather simplistic, often focusing upon incremental research challenges. At the same time, many of the recent language resources were annotated on a rather shallow level, with only a few highly generalised and abstract labels, often using crowd-workers who are only able to produce large amounts of consistent and high quality annotations if the annotation task is rather simple and does not require expert linguistic knowledge or in-depth training (Poesio et al. 2017, call these "microtasks"). In short, since the neural turn in approx. 2014/2015 we can observe a trend towards *simply more and more* annotations with increasing quantity while ignoring complexity and structure, and also a trend towards *more and more simple* annotations that are cheaper to produce and easier to generalise from. Has annotation science perhaps become obsolete? Have the lessons and insights learned in the last 30 years become irrelevant, given today's popularity and power of neural approaches for processing and, perhaps, finally, understanding language?

Acknowledgment: This chapter is based on a presentation given at the conference *Annotation in Scholarly Editions and Research: Function – Differentiation – Systematization*, held at the University of Wuppertal, Germany, on 20–22 February 2019. The author would like to thank the organisers, Julia Nantke and Frederik Schlupkothen, for the invitation and especially for their patience. Peter Bourgonje and Karolina Victoria Zaczynska provided comments on an early draft of this article for

which the author is grateful. Work on this chapter was partially supported by the projects ELG (EU Horizon 2020, no. 825627), LYNX (EU Horizon 2020, no. 780602) and QURATOR (BMBF, no. 03WKDA1A).

Bibliography

Artstein, Ron. Inter-annotator Agreement. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 297–313.

Bayerl, Petra Saskia and Karsten Ingmar Paul. Squibs and Discussions: Identifying Sources of Disagreement: Generalizability Theory in Manual Annotation Studies. In: Computational Linguistics, 33(1). March 2007, pp. 3–8. DOI: 10.1162/coli.2007.33.1.3.

Bayerl, Petra Saskia and Karsten Ingmar Paul. What Determines Inter-Coder Agreement in Manual Annotations? A Meta-Analytic Investigation. In: Computational Linguistics, 37(4). Cambridge, MA: The MIT Press Journals. December 2011, pp. 699–725. DOI: 10.1162/COLI_a_00074.

Berners-Lee, Tim. Information Management: A Proposal. 1989. URL: https://cds.cern.ch/record/369245/files/dd-89-001.pdf (04.02.2020).

Biemann, Chris, Kalina Bontcheva, Richard Eckart de Castilho, Iryna Gurevych, and Seid Muhie Yimam. Collaborative Web-Based Tools for Multi-layer Text Annotation. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 229–256.

Bird, Steven and Mark Liberman. A Formal Framework for Linguistic Annotation. In: Speech Communication, 33,1–2. Elsevier. January 2001, pp. 23–60. DOI: 10.1016/S0167-6393(00)00068-6.

Bludau, Mark-Jan, Marian Dörk, Heiner Fangerau, Thorsten Halling, Elena Leitner, Sina Menzel, Gerhard Müller, Vivien Petras, Georg Rehm, Clemens Neudecker, David Zellhöfer, and Julián Moreno Schneider. SoNAR (IDH): Datenschnittstellen für historische Netzwerkanalyse. In: Christof Schöch (Ed.), DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts. Paderborn, Germany. March 2020, pp. 360–362. DOI: 10.5281/zenodo.3666690.

Bourgonje, Peter, Julián Moreno-Schneider, Jan Nehring, Georg Rehm, Felix Sasaki, and Ankit Srivastava. Towards a Platform for Curation Technologies: Enriching Text Collections with a Semantic-Web Layer. In: Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, Dunja Mladenić, Sören Auer, and Christoph Lange (Eds.), The Semantic Web, number 9989 in Lecture Notes in Computer Science: ESWC 2016 Satellite Events. Heraklion, Crete, Greece, May 29 – June 2, 2016. Revised Selected Papers. Cham, Switzerland: Springer. June 2016, pp. 65–68.

Bray, Tim, Jean Paoli, C. Michael Sperberg-McQueen, Eve Maler, and François Yergeau. Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation, World Wide Web Consortium (W3C). November 2008. URL: https://www.w3.org/TR/xml/ (04.02.2020).

Bush, Vannevar. As we may think. In: Atlantic Monthly, 176. Fortune Magazine. 1945, pp. 101–108.

Chiarcos, Christian and Maria Sukhareva. OLiA – Ontologies of Linguistic Annotation. In: Semantic Web Journal, 6(4). Clifton, VA, Amsterdam: IOS Press. 2015, pp. 379–386. DOI: 10.3233/SW-140167.

Dickinson, Markus and Dan Tufiş. Iterative Enhancement. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 257–276.

Dipper, Stefanie, Michael Götze, and Manfred Stede. Simple Annotation Tools for Complex Annotation Tasks: An Evaluation. In: Proceedings of the LREC Workshop on XML-based richly annotated corpora. May 2004, pp. 54–62. URL: https://www.linguistics.rub.de/~dipper/pub/xbrac04-sfb.pdf (04.02.2020).

Eckart de Castilho, Richard, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In: Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH). Osaka, Japan. December 2016, pp. 76–84. URL: https://www.aclweb.org/anthology/W16-4011.pdf (04.02.2020).

Farrar, Scott and D. Terrence Langendoen. An OWL-DL Implementation of Gold: An Ontology for the Semantic Web. In: Dieter Metzing and Andreas Witt (Eds.), Linguistic Modeling of Information and Markup Languages. Contributions to Language Technology. Dordrecht: Springer. 2010, pp. 45–66.

Finlayson, Mark A. and Tomaž Erjavec. Overview of Annotation Creation: Processes and Tools. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 167–191.

Goecke, Daniela, Harald Lüngen, Dieter Metzing, Maik Stührenberg, and Andreas Witt. Different Views on Markup – Distinguishing Levels and Layers. In: Dieter Metzing and Andreas Witt (Eds.), Linguistic Modeling of Information and Markup Languages. Contributions to Language Technology. Dordrecht: Springer. 2010, pp. 1–21.

Gut, Ulrike and Petra Saskia Bayerl. Measuring the Reliability of Manual Annotations of Speech Corpora. In: Proceedings of the 2nd International Conference on Speech Prosody. Nara, Japan: ISCA Archive. January 2004, pp. 565–568. URL: https://www.isca-speech.org/archive_open/sp2004/sp04_565.pdf (04.02.2020).

Hellmann, Sebastian, Jens Lehmann, Sören Auer, and Martin Brümmer. Integrating NLP Using Linked Data. In: Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha Noy, Chris Welty, and Krzysztof Janowicz (Eds.), The Semantic Web – Proceedings of ISWC. Sydney, Australia. Berlin, Heidelberg: Springer. 21–25 October 2013, pp. 98–113.

Ide, Nancy. Annotation Science: From Theory to Practice and Use. In: Georg Rehm, Andreas Witt, and Lothar Lemnitzer (Eds.), Datenstrukturen für linguistische Ressourcen und ihre Anwendungen – Data Structures for Linguistic Resources and Applications: Proceedings of the Biennial GLDV Conference. Tübingen: Gunter Narr. 2007, pp. 3–7.

Ide, Nancy and James Pustejovsky (Eds.). Handbook of Linguistic Annotation. Dordrecht: Springer. 2017.

Ide, Nancy and Laurent Romary. A Common Framework for Syntactic Annotation. In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, ACL'01. Stroudsburg, PA: Association for Computational Linguistics. July 2001, pp. 306–313. DOI: 10.3115/1073012.1073052.

Ide, Nancy and Keith Suderman. GrAF: A Graph-based Format for Linguistic Annotations. In: Proceedings of the Linguistic Annotation Workshop. Prague, Czech Republic: Association for Computational Linguistics. June 2007, pp. 1–8. URL: https://www.aclweb.org/anthology/W07-1501.pdf (04.02.2020).

Ide, Nancy, Nicoletta Calzolari, Judith Eckle-Kohler, Dafydd Gibbon, Sebastian Hellmann, Kiyong Lee, Joakim Nivre, and Laurent Romary. Community Standards for Linguistically-Annotated

Resources. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017a, pp. 113–165.

Ide, Nancy, Christian Chiarcos, Manfred Stede, and Steve Cassidy. Designing Annotation Schemes: From Model to Representation. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017b, pp. 73–111.

Jaradeh, Mohamad Yaser, Allard Oelen, Kheir Eddine Farfar, Manuel Prinz, Jennifer D'Souza, Gábor Kismihók, Markus Stocker, and Sören Auer. Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture, K-CAP '19. New York, NY: ACM. 2019, pp. 243–246. DOI: 10.1145/3360901.3364435.

Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. Association for Computational Linguistics. June 2018, pp. 5–9. URL: https://www.aclweb.org/anthology/C18-2002.pdf (04.02.2020).

Labropoulou, Penny, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, Maria Berger, Valérie Mapelli, Michael Rigault, Victoria Arranz, Khalid Choukri, Gerhard Backfried, José Manuel Gómez Pérez, and Andres Garcia-Silva. Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid. In: Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.), Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), Marseille, France: European Language Resources Association (ELRA). 2020. Accepted for publication.

Leitner, Elena, Georg Rehm, and Julián Moreno-Schneider. Fine-grained Named Entity Recognition in Legal Documents. In: Maribel Acosta, Philippe Cudré-Mauroux, Maria Maleshkova, Tassilo Pellegrini, Harald Sack, and York Sure-Vetter (Eds.), Semantic Systems. The Power of AI and Knowledge Graphs. Proceedings of the 15th International Conference (SEMANTiCS 2019), number 11702 in Lecture Notes in Computer Science. Karlsruhe, Germany: Springer. 10–11 September 2019, pp. 272–287.

Leitner, Elena, Georg Rehm, and Julián Moreno-Schneider. A Dataset of German Legal Documents for Named Entity Recognition. In: Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.), Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), Marseille, France: European Language Resources Association (ELRA). 2020. Accepted for publication.

Maler, Eve and Jeanne El Andaloussi. Developing SGML DTDs – From Text to Model to Markup. Upper Saddle River, New Jersey: Prentice Hall. 1996.

Megginson, Dave. Structuring XML Documents. Charles F. Goldfarb Series on Open Information Management. Upper Saddle River, NJ: Prentice Hal. 1998.

Meister, Jan Christoph, Jan Horstmann, Marco Petris, Janina Jacke, Christian Bruck, Mareike Schumacher, and Marie Flüh. CATMA – Computer Assisted Text Markup and Analysis. 2019. URL: https://catma.de/ (04.02.2020).

Metzing, Dieter and Andreas Witt (Eds.). Linguistic Modelling of Information and Markup Languages. Contributions to Language Technology. Dordrecht, Heidelberg, London, New York: Springer. 2010.

Moreno-Schneider, Julián, Ankit Srivastava, Peter Bourgonje, David Wabnitz, and Georg Rehm. Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard. In: Octavian Popescu and Carlo Strapparava (Eds.), Proceedings of the Second Workshop on Natural Language Processing meets Journalism – EMNLP 2017 Workshop (NLPMJ 2017). Copenhagen, Denmark. 2017, pp. 68–73. URL: https://www.aclweb.org/anthology/W17-4212.pdf (04.02.2020).

Nelson, Theodor Holm. Literary Machines. Edition 87.1. Sausalito, CA: Mindful Press. 1987.

Neves, Mariana and Jurica Ševa. An Extensive Review of Tools for Manual Annotation of Documents. In: Briefings in Bioinformatics. Oxford: Oxford Academic. 2019. DOI: 10.1093/bib/bbz130.

Ostendorff, Malte, Peter Bourgonje, Maria Berger, Julián Moreno-Schneider, and Georg Rehm. Enriching BERT with Knowledge Graph Embeddings for Document Classification. In: Steffen Remus, Rami Aly, and Chris Biemann (Eds.), Proceedings of the GermEval Workshop 2019 – Shared Task on the Hierarchical Classification of Blurbs. Erlangen, Germany. October 2019.

Piperidis, Stelios, Harris Papageorgiou, Christian Spurk, Georg Rehm, Khalid Choukri, Olivier Hamon, Nicoletta Calzolari, Riccardo del Gratta, Bernardo Magnini, and Christian Girardi. META-SHARE: One year after. In: Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.), Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014). Reykjavik, Iceland: European Language Resources Association (ELRA). May 2014, pp. 1532–1538.

Poesio, Massimo, Jon Chamberlain, and Udo Kruschwitz. Crowdsourcing. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 277–295. DOI: 10.1007/978-94-024-0881-2_10.

Pustejovsky, James, Harry Bunt, and Annie Zaenen. Designing Annotation Schemes: From Theory to Model. In: Nancy Ide and James Pustejovsky (Eds.), Handbook of Linguistic Annotation. Dordrecht: Springer. 2017, pp. 21–72.

Rehm, Georg. Towards Automatic Web Genre Identification – A Corpus-Based Approach in the Domain of Academia by Example of the Academic's Personal Homepage. In: Ralph Sprague (Ed.), Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35). Big Island, Hawaii: IEEE Computer Society. January 2002, pp. 1143–1152.

Rehm, Georg. Hypertextsorten: Definition – Struktur – Klassifikation. Norderstedt: Books on Demand. 2007. PhD thesis in Applied and Computational Linguistics, Justus-Liebig-Universität Gießen, 2005.

Rehm, Georg. Hypertext Types and Markup Languages – The Relationship Between HTML and Web Genres. In: Dieter Metzing and Andreas Witt (Eds.), Linguistic Modeling of Information and Markup Languages. Contributions to Language Technology. Dordrecht: Springer. 2010a, pp. 143–164.

Rehm, Georg. Texttechnologische Grundlagen. In: Kai-Uwe Carstensen, Christian Ebert, Cornelia Endriss, Susanne Jekat, Ralf Klabunde, and Hagen Langer (Eds.), Computerlinguistik und Sprachtechnologie – Eine Einführung 3. edition. Heidelberg: Spektrum. 2010b, pp. 159–168.

Rehm, Georg. The Language Resource Life Cycle: Towards a Generic Model for Creating, Maintaining, Using and Distributing Language Resources. In: Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.), Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016). Portorož, Slovenia: European Language Resources Association (ELRA). May 2016, pp. 2450–2454.

Rehm, Georg. An Infrastructure for Empowering Internet Users to handle Fake News and other Online Media Phenomena. In: Georg Rehm and Thierry Declerck (Eds.), Language Technologies for the Challenges of the Digital Age: 27th International Conference, GSCL 2017. Berlin, Germany. 13–14 September 2017. Proceedings, number 10713 in Lecture Notes in Artificial Intelligence (LNAI). Cham, Switzerland: Springer. January 2018, pp. 216–231.

Rehm, Georg, Richard Eckart, and Christian Chiarcos. An OWL- and XQuery-Based Mechanism for the Retrieval of Linguistic Patterns from XML-Corpora. In: Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, Nicolas Nicolov, and Nicolai Nikolov (Eds.), International Conference Recent Advances in Natural Language Processing (RANLP 2007). Borovets, Bulgaria: Incoma. September 2007a, pp 510–514.

Rehm, Georg, Andreas Witt, and Lothar Lemnitzer (Eds.). Datenstrukturen für linguistische Ressourcen und ihre Anwendungen – Data Structures for Linguistic Resources and Applications: Proceedings of the Biennial GLDV Conference 2007. Tübingen: Gunter Narr. 2007b.

Rehm, Georg, Richard Eckart, Christian Chiarcos, and Johannes Dellert. Ontology-Based XQuery'ing of XML-Encoded Language Resources on Multiple Annotation Layers. In: Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias (Eds.), Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008). Marrakesh, Morocco: European Language Resources Association. May 2008a, pp. 525–532. URL: https://www.cs.brandeis.edu/~marc/misc/proceedings/lrec-2008/pdf/139_paper.pdf (04.02.2020).

Rehm, Georg, Oliver Schonefeld, Andreas Witt, Timm Lehmberg, Christian Chiarcos, Hanan Bechara, Florian Eishold, Kilian Evang, Magdalena Leshtanska, Aleksandar Savkov, and Matthias Stark. The Metadata-Database of a Next Generation Sustainability Web- Platform for Language Resources. In: Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias (Eds.), Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008). Marrakesh, Morocco: European Language Resources Association. May 2008b, pp. 371–378. URL: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4508/file/Rehm_Schonefeld_Witt_The_Meta_data_Database_of_a_Next_Generation_2008.pdf (04.02.2020).

Rehm, Georg, Oliver Schonefeld, Andreas Witt, Erhard Hinrichs, and Marga Reis. Sustainability of Annotated Resources in Linguistics: A Web-Platform for Exploring, Querying and Distributing Linguistic Corpora and Other Resources. In: Literary and Linguistic Computing, 24(2). Oxford: Oxford University Press. 2009, pp. 193–210. DOI: 10.1093/llc/fqp003.

Rehm, Georg, Felix Sasaki, and Aljoscha Burchardt. Web Annotations – A Game Changer for Language Technologies? I Annotate 2016. Berlin, Germany. May 2016. URL: https://de.slideshare.net/georgrehm/web-annotations-a-game-changer-for-language-technology (04.02.2020).

Rehm, Georg, Julián Moreno Schneider, and Peter Bourgonje. Automatic and Manual Web Annotations in an Infrastructure to handle Fake News and other Online Media Phenomena. In: Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga (Eds.), Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA). May 2018a, pp. 2416–2422.

Rehm, Georg, Julián Moreno-Schneider, Peter Bourgonje, Ankit Srivastava, Rolf Fricke, Jan Thomsen, Jing He, Joachim Quantz, Armin Berger, Luca König, Sören Räuchle, Jens Gerth, and

David Wabnitz. Different Types of Automated and Semi-Automated Semantic Storytelling: Curation Technologies for Different Sectors. In: Georg Rehm and Thierry Declerck (Eds.), Language Technologies for the Challenges of the Digital Age: 27th International Conference, GSCL 2017. Berlin, Germany. 13.–14. September 2017. Proceedings, number 10713 in Lecture Notes in Artificial Intelligence (LNAI). Cham, Switzerland: Gesellschaft für Sprachtechnologie und Computerlinguistik e.V., Springer. January 2018b, pp. 232–247.

Rehm, Georg, Julián Moreno-Schneider, Jorge Gracia, Artem Revenko, Victor Mireles, Maria Khvalchik, Ilan Kernerman, Andis Lagzdins, Marcis Pinnis, Artus Vasilevskis, Elena Leitner, Jan Milde, and Pia Weißenhorn. Developing and Orchestrating a Portfolio of Natural Legal Language Processing and Document Curation Services. In: Nikolaos Aletras, Elliott Ash, Leslie Barrett, Daniel Chen, Adam Meyers, Daniel Preotiuc-Pietro, David Rosenberg, and Amanda Stent (Eds.), Proceedings of Workshop on Natural Legal Language Processing (NLLP 2019). Minneapolis, USA. June 2019, pp. 55–66. URL: https://www.aclweb.org/anthology/W19-2207.pdf (04.02.2020).

Rehm, Georg, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitrios Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajic, Jana Hamrlova, Lukas Kacena, Khalid Choukri, Victoria Arranz, Valerie Mapelli, Andrejs Vasiljevs, Orians Anvari, Andis Lagzdins, Julija Melnika, Gerhard Backfried, Erinç Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampler, Dorothea Thomas-Aniola, Jose M. Gomez Perez, Andres Garcia Silva, Cristian Berrio, Ulrich Germann, Steve Renals, and Ondrej Klejch. European Language Grid: An Overview. 2020a. Submitted to LREC 2020. URL: https://www.european-language-grid.eu/wp-content/uploads/2019/10/00-03-ELG-Overview-Georg-Rehm.pdf (04.02.2020).

Rehm, Georg, Peter Bourgonje, Stefanie Hegele, Florian Kintzel, Julián Moreno-Schneider, Malte Ostendorff, Karolina Zaczynska, Armin Berger, Stefan Grill, Sören Räuchle, Jens Rauenbusch, Lisa Rutenburg, André Schmidt, Mikka Wild, Henry Hoffmann, Julian Fink, Sarah Schulz, Jurica Seva, Joachim Quantz, Joachim Böttger, Josefine Matthey, Rolf Fricke, Jan Thomsen, Adrian Paschke, Jamal Al Qundus, Thomas Hoppe, Naouel Karam, Frauke Weichhardt, Christian Fillies, Clemens Neudecker, Mike Gerber, Kai Labusch, Vahid Rezanezhad, Robin Schaefer, David Zellhöfer, Daniel Siewert, Patrick Bunk, Lydia Pintscher, Elena Aleynikova, and Franziska Heine. QURATOR: Innovative Technologies for Content and Data Curation. In: Adrian Paschke, Clemens Neudecker, Georg Rehm, Jamal Al Qundus, and Lydia Pintscher (Eds.), Proceedings of QURATOR 2020 – The conference for intelligent content solutions. Berin, Germany. February 2020. CEUR Workshop Proceedings, Volume 2535. 20.–21. January 2020b.

Rehm, Georg, Dimitrios Galanis, Penny Labropoulou, Stelios Piperidis, Martin Welß, Ricardo Usbeck, Joachim Köhler, Miltos Deligiannis, Katerina Gkirtzou, Johannes Fischer, Christian Chiarcos, Nils Feldhus, Julián Moreno-Schneider, Florian Kintzel, Elena Montiel, Víctor Rodríguez Doncel, John P. McCrae, David Laqua, Irina Patricia Theile, Christian Dittmar, Kalina Bontcheva, Ian Roberts, Andrejs Vasiljevs, and Andis Lagzdiņš. Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability. In: Georg Rehm, Kalina Bontcheva, Khalid Choukri, Jan Hajic, Stelios Piperidis, and Andrejs Vasiljevs (Eds.), Proceedings of the 1st International Workshop on Language Technology Platforms (IWLTP 2020, co-located with LREC 2020). Marseille, France. May 2020c, pp. 96–107.

Rehm, Georg, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Christoph Prinz, José Manuel

Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Albina Auksoriūtė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabík, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiș, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon. The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe. In: Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.), Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). Marseille, France: European Language Resources Association (ELRA). 2020d. Accepted for publication.

Sanderson, Robert. Web Annotation Protocol. W3C Recommendation. World Wide Web Consortium (W3C). February 2017a. URL: https://www.w3.org/TR/annotation-protocol/ (04.02.2020).

Sanderson, Robert, Paolo Ciccarese, and Benjamin Young. Web Annotation Data Model. W3C Recommendation. World Wide Web Consortium (W3C). February 2017b. URL: https://www.w3.org/TR/annotation-model/ (04.02.2020).

Sanderson, Robert, Paolo Ciccarese, and Benjamin Young. Web Annotation Vocabulary. W3C Recommendation. World Wide Web Consortium (W3C). February 2017c. URL: https://www.w3.org/TR/annotation-vocab/ (04.20.2020).

Snow, Rion, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. Cheap and Fast—But is it Good?: Evaluating Non-Expert Annotations For Natural Language Tasks. In: Proceedings of the conference on empirical methods in natural language processing. Honolulu, Hawaii: Association for Computational Linguistics. October 2008, pp. 254–263.

Stede, Manfred. Korpusgestützte Textanalyse: Grundzüge der Ebenen-orientierten Textlinguistik. 2. überarbeitete Auflage. Tübingen: Narr Francke Attempto. 2018.

Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a Web-based Tool for NLP-Assisted Text Annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France: Association for Computational Linguistics. April 2012, pp. 102–107.

Stührenberg, Maik. The TEI and current standards for structuring linguistic data. An overview. In: Journal of the Text Encoding Initiative, 3. November 2012. DOI: 10.4000/jtei.523.

TEI Consortium (Ed.). TEI: P5 Guidelines for Electronic Text Encoding and Interchange. TEI Consortium, 2019. Version 3.6.0. Last updated on 16th July 2019. Originally edited by C. M. Sperberg-McQueen and Lou Burnard for the ACH-ALLC-ACL Text Encoding Initiative, now entirely revised and expanded under the supervision of the Technical Council of the TEI Consortium. URL: https://tei-c.org/Vault/P5/3.6.0/doc/tei-p5-doc/en/html/ (04.02.2020).

Walsh, Norman and Tovey Bethan. The Markup Declaration. In: B. Tommie Usdin (Ed.), Proceedings of Balisage: The Markup Conference 2018, Washington, DC. Balisage Series on Markup Technologies, Vol. 21. 2018. DOI: 10.4242/BalisageVol21.Tovey01.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair

J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for Scientific Data Management and Stewardship. In: Scientific Data, 3. 2016. DOI: 10.1038/sdata.2016.18.

Witt, Andreas, Georg Rehm, Timm Lehmberg, and Erhard Hinrichs. Mapping Multi-Rooted Trees from a Sustainable Exchange Format to TEI Feature Structures. In: TEI@20: 20 Years of Supporting the Digital Humanities. The 20th Anniversary Text Encoding Initiative Consortium Members' Meeting, University of Maryland, College Park, 10. 2007a. URL: https://www.tei-c.org/Vault/MembersMeetings/2007/ (04.02.2020).

Witt, Andreas, Oliver Schonefeld, Georg Rehm, Jonathan Khoo, and Kilian Evang. On the Lossless Transformation of Single-File, Multi-Layer Annotations into Multi-Rooted Trees. In: B. Tommie Usdin (Ed.), Proceedings of Extreme Markup Languages 2007. Montréal, Canada. August 2007b. URL: http://conferences.idealliance.org/extreme/html/2007/Witt01/EML2007Witt01.xml (04.02.2020).

Witt, Andreas, Georg Rehm, Erhard Hinrichs, Timm Lehmberg, and Jens Stegmann. SusTEInability of Linguistic Resources through Feature Structures. In: Literary and Linguistic Computing, 24(3). Oxford: Oxford University Press. 2009, pp. 363–372.

Wörner, Kai, Andreas Witt, Georg Rehm, and Stefanie Dipper. Modelling Linguistic Data Structures. In: B. Tommie Usdin (Ed.), Proceedings of Extreme Markup Languages 2006. Montréal, Canada. August 2006. URL: http://conferences.idealliance.org/extreme/html/2006/Witt01/EML2006Witt01.html (04.02.2020).