Published in: Borin, Lars (ed.): Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure. - Linköping: Linköping University Electronic Press, 2017. S. 113-130 (Linköping Electronic Conference Proceedings)

Conversion and Annotation Web Services for Spoken Language Data in CLARIN

Thomas Schmidt
Institute for the German Language (IDS)
Mannheim
Germany
thomas.schmidt@ids-mannheim.de

Hanna Hedeland
Hamburg Centre for Language Corpora (HZSK)
University of Hamburg
Germany
hanna.hedeland@uni-hamburg.de

Daniel Jettka
Hamburg Centre for Language Corpora (HZSK)
University of Hamburg
Germany
daniel.jettka@uni-hamburg.de

Abstract

We present an approach to making existing CLARIN web services usable for spoken language transcriptions. Our approach is based on a new TEI-based ISO standard for such transcriptions. We show how existing tool formats can be transformed to this standard, how an encoder/decoder pair for the TCF format enables users to feed this type of data through a WebLicht tool chain, and why and how web services operating directly on the standard format would be useful.

1 Introduction

Web services operating on language resources are a central idea for the CLARIN infrastructure. This includes services for the annotation of text data, such as lemmatizers, Part-Of-Speech-Taggers, Named Entity Recognizers, and so forth. WebLicht (Hinrichs et al., 2010) is an application that integrates many such services into a common web-based framework and enables users to build and apply annotation chains for a given set of data. So far, most such services were built with, and are meant to operate on, "canonical" written language data, typically edited texts from newspapers, books, etc., in the standard orthography of a major language. For "non-canonical" types of written data (such as CMC data[1], or historical language data[2]) and for spoken language data, these services are often not directly usable for at least two reasons:

a. The data come in formats which are more complex than the simple "stream of tokens" (see Menke et al., 2015) expected by many annotation services. They can contain transcriptions of simultaneous ("overlapping") speech or two or more alternative transcriptions for the same stretch of speech (if the transcriber is uncertain), both of which require parallel structures to be encoded in the data format.

b. The text data may require additional processing steps or adaptations of annotation methods in order to yield useful results. Not all "tokens" of a spoken language transcription are simple words or punctuation. We also find descriptions of pauses or non-verbal actions, which may have to undergo additional processing before they can be fed into, say, a Part-Of-Speech-Tagger. The use of non-standardized writing (as in "modified orthography" to represent pronunciation deviating from the standard), the lack or diverging use of punctuation (as in systems which use punctuation to represent prosodic properties of speech), semi-lexical material (like hesitation markers or interjections), or incomplete tokens (as in a repair sequence or an aborted utterance) may cause similar problems.

[1] See: http://de.clarin.eu/en/curation-project-1-3-german-philology.
[2] See: http://www.deutschestextarchiv.de/doku/software.

Thomas Schmidt, Hanna Hedeland and Daniel Jettka 2017. Conversion and annotation web services for spoken language data in CLARIN. *Selected papers from the CLARIN Annual Conference 2016*. Linköping Electronic Conference Proceedings 136: 113–130.

We present an approach to making existing services and service environments in CLARIN usable for spoken language data. That this is possible and useful in principle has already been demonstrated by proof-of-concept implementations in the tools ELAN (Sloetjes, 2014) and EXMARaLDA (Schmidt and Wörner, 2014, see section 6), which both provide interfaces from their respective tool formats to WebLicht and WebMaus (Kisler et al., 2012). The approach described here is potentially more flexible because it is based on a recently published TEI-based ISO standard for spoken language transcriptions and thus not directly tied to a specific tool or tool format.

The general architecture we envision and have started to implement is depicted in Figure 1. The point of departure for most users will be one of a few established formats of tools for multimedia annotation. This needs to be converted to the ISO/TEI standard format which then constitutes the basis for all further processing steps. Existing services in WebLicht can be made usable by providing an encoder/decoder pair to/from WebLicht's TCF format. Additional services specializing on spoken language data can operate directly on the ISO/TEI data.

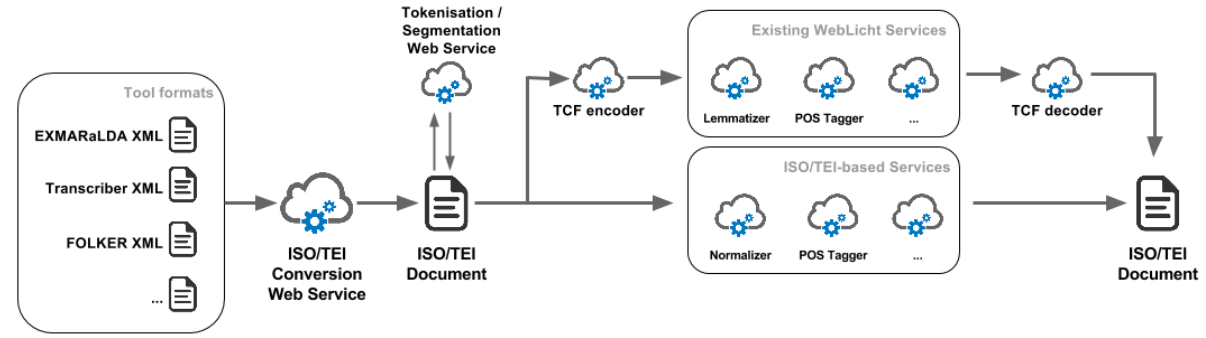

Figure 1: Architecture

We briefly sketch the most important characteristics of the ISO/TEI format in section 2. Sections 3 (conversion from tool to ISO/TEI) and 4 (TCF encoder/decoder) describe the conversion steps needed to connect existing tools with WebLicht via the ISO standard. Section 5 addresses some details of the CLARIN integration of these services, while section 6 deals with the way users can interact with them. In section 7, we briefly introduce the idea of web services directly operating on the ISO/TEI standard.

2 A TEI-Based ISO Standard for Spoken Language Transcriptions

The Text Encoding Initiative's Guidelines (TEI, chapter 8)[3], have always contained suggestions for representing spoken language transcriptions within the TEI framework. However, this portion of the guidelines has never been sufficiently established in the respective research communities to really work as an interchange format or even a standard. As Bird and Liberman (2001, p. 26) put it:

> The TEI guidelines for 'Transcriptions of Speech' offer access to a very broad range of representational techniques drawn from other aspects of the TEI specification. The TEI report sketches or alludes to a correspondingly wide range of possible issues in speech annotation. All of these seem to be encompassed within [the author's AG framework], but it does not seem appropriate to speculate at much greater length about this, given that this portion of the TEI guidelines does not seem to have been used in any published transcriptions to date. [emphasis added]

One reason for this lack of acceptance is that work with spoken language data relies on specialized tools for efficient transcription, and the relation of TEI to these tools was never sufficiently clarified. Schmidt (2005) therefore made a first suggestion on how to reconcile the time-based data models, which most tools are based on, with the hierarchy-based data model underlying the TEI proposal (see also Parisse and Morgenstern, 2010, for a similar approach). Taking into account findings from a tool interoperability study (Schmidt et al., 2009), the proposal for representing spoken language transcriptions on the basis of TEI was further refined in Schmidt (2011), and from 2012 on, became the topic of an ISO standardization project (ISO/ TC 37/SC 4/WG 6), concluded in summer 2016 with the official publication of "Language resource management - Transcription of Spoken Language (24642)" (see ISO,

[3] See: http://www.tei-c.org/Vault/P5/1.0.0/doc/tei-p5-doc/en/html/TS.html.

2016). We will briefly outline some characteristics of this standard that are important to the subject of this paper.

The general focus of the standard is on orthography-based (i.e. not: IPA-based) transcription of recordings of authentic interaction (i.e. not: monologic "speech" or experiment data). Guiding design principles were the maxim to reuse as many elements as possible from chapter 8 of the existing TEI guidelines[4] and to orient their use towards interoperability with established tools. In particular, this meant a conscious limitation of choices in the numerous cases where the guidelines offer more than one concept for representing one and the same phenomenon.

As exemplified in Figure 2, basic building blocks for the document structure are (a) one or more <recording> elements specifying the underlying audio and/or video file(s), (b) a <particDesc> defining the participants of the interaction, and (c) a <timeline> providing offsets into a recording.

```
(a) <sourceDesc>
       <recordingStmt>
          <recording type="video">
            <media mimeType="audio/wav"
                url="file:/corpus/media/interaction_101.wav"/>
          </recording>
       </recordingStmt>
    </sourceDesc>

(b)  <profileDesc>
        <particDesc>
           <person xml:id="MJ" n="Mick" sex="1"/>
           <person xml:id="KR" n="Keith" sex="1"/>
        </particDesc>
        <!-- [...] -->
     </profileDesc>


(c)  <timeline unit="s" origin="#T0">
        <when xml:id="T0"/>
        <when xml:id="T1" interval="0.906636353362215" since="#T0"/>
        <when xml:id="T2" interval="2.6012168690059636" since="#T0"/>
        <!-- [...] -->
    </timeline>
```

Figure 2: Recording(s), participant(s) and the timeline as basic building blocks

The main part of the document is then made up of a sequence of <u> elements. They correspond to individual speaker contributions and contain the actual transcription text, references to points in the timeline and to the respective speaker (@who). As illustrated in Figure 3, the standard allows different levels of detail for the actual markup of the transcription text. In the simplest case (example 1 in Figure 3), a plain text string can be used, which is temporally aligned via mandatory @start and @end attributes of the <u> element. Intervening temporal alignment can be added in the form of additional <anchor> milestone elements (example 2) whose @synch attribute refers to a point in the timeline.

The microstructure of the speaker contribution can be represented by inserting additional markup, most importantly <w> for word tokens, <pause> for pauses and <vocal> or <kinesic> for non-verbal phenomena like coughing or laughing (example 3). Finally, segmentations of speaker contributions into units above the word level (the "sentence equivalents" of spoken language such as intonation units) can be represented by intervening <seg> elements (example 4). As we will discuss below, the additional markup below <u> is crucial for many automatic annotation methods. The examples in Figure 3 illustrate transcription proper, i.e. the direct representation in written form of what is heard or seen in the primary data.

[4] See: http://www.tei-c.org/Vault/P5/1.0.0/doc/tei-p5-doc/en/html/TS.html.

```
(1) <u who="MJ" start="#T0" end="#T2">
        I ((cough)) see a door. I (0.3) want to paint it (black/blue).
    </u>

(2) <u who="MJ" start="#T0" end="#T2">
        I ((cough)) see a door.
        <anchor synch="#T1"/>
        I (0.3) want to paint it (black/blue).
     </u>
(3) <u who="MJ" start="#T0" end="#T2">
        <w>I</w>
        <vocal><desc>cough</desc></vocal>
        <w>see</w><w>a</w><w>door</w><p>.</p>
        <anchor synch="#T1"/>
        <w>I/w>
        <pause dur="PT0.3S"/>
        <w>want</w><w>to</w><w>paint</w><w>it/w>
        <unclear><choice><w>black</w><w>blue</w></choice></unclear>
        <p>.</p>
     </u>

(4) <u who="MJ" start="#T0" end="#T2">
        <seg type="intonation-phrase" subtype="falling">
                <w>I</w>
                <vocal><desc>cough</desc></vocal>
                <w>see</w><w>a</w><w>door</w>
        </seg>
        <anchor synch="#T1"/>
        <seg type="intonation-phrase" subtype="falling">
                <w>I/w>
                <pause dur="PT0.3S"/>
                <w>want</w><w>to</w><w>paint</w><w>it/w>
                <unclear><choice><w>black</w><w>blue</w></choice></unclear>
        </seg>
    </u>
```

Figure 3: <u> elements with different levels of internal markup

In order to represent additional annotations on that material, the ISO/TEI standard provides the possibility to introduce an arbitrary number of standoff annotation layers in <spanGrp> elements and to group these with the <u> element they belong to, using an <annotationBlock> element (see Banski et al., 2016). This mechanism is crucial also for storing annotations that result from automatic annotation methods in WebLicht. Figure 4 illustrates the annotation of an utterance with lemmas and part-of-speech tags.

```
<annotationBlock who="MJ" start="#T0" end="#T2" xml:id="ab1">
   <u xml:id="u1">
        <seg type="intonation-phrase" subtype="falling" xml:id="seg1">
                <w xml:id="w1">I</w>
                <vocal xml:id="voc1"><desc>cough</desc></vocal>
                <w xml:id="w2">see</w>
                <w xml:id="w3">a</w>
                <w xml:id="w4">door</w>
        </seg>
   </u>
   <spanGrp type="lemma">
        <span from="#w1" to="#w1">I</span>
        <span from="#w2" to="#w2">see</span>
        <span from="#w3" to="#w3">a</span>
        <span from="#w4" to="#w4">door</span>
   </spanGrp>
   <spanGrp type="pos">
        <span from="#w1" to="#w1">PPER</span>
        <span from="#w2" to="#w2">V</span>
        <span from="#w3" to="#w3">DET</span>
        <span from="#w4" to="#w4">NN</span>
   </spanGrp>
</annotationBlock>
```

Figure 4: <annotationBlock> grouping an <u> with standoff annotation (lemmatization and POS tagging) in <spanGrp>

With the same mechanism (figure 5), orthographically normalized forms can be assigned to transcribed forms when the latter do not follow standard orthography. This is the case, for instance, in many conversation analytic transcription systems that use "literary transcription" or "eye dialect" to represent actual pronunciations that deviate from the ones suggested by the orthographic form (as in "gotta" for "got to").

```
<annotationBlock who="CB" start="#T0" end="#T2" xml:id="ab1">
   <u xml:id="u1">
        <w xml:id="w1">sure</w>
        <w xml:id="w2">nuff</w>
        <w xml:id="w3">an</w>
        <w xml:id="w4">yes</w>
        <w xml:id="w5">I</w>
        <w xml:id="w6">do</w>
   </u>
   <spanGrp type="normalized">
        <span from="#w1" to="#w1">sure</span>
        <span from="#w2" to="#w2">enough</span>
        <span from="#w3" to="#w3">and</span>
        <span from="#w4" to="#w4">yes</span>
        <span from="#w5" to="#w5">I</span>
        <span from="#w6" to="#w6">do</span>
   </spanGrp>
</annotationBlock>
```

Figure 5: <annotationBlock> grouping an <u> with standoff annotation (orthographically normalized forms) in <spanGrp>

For the spoken language data hosted at the CLARIN centers in Hamburg (HZSK) and Mannheim (IDS/AGD), we have confirmed that a lossless, fully automatic transformation from existing formats (mostly EXMARaLDA and/or FOLKER) to ISO compliant TEI is possible. Current and future developments at these data centers will make this standard a central element of all workflows.

In the context of the Parthenos project[5], support will be given via INRIA to further develop and document the standard and disseminate it to the scientific community.

3 Converting Common Transcription Formats to ISO/TEI

Unlike other text types (such as manuscripts, dictionaries etc.) addressed by the TEI, spoken language transcription is rarely done by editing an XML document directly. Researchers crucially rely on tools which support the alignment of sound/video and transcription in an ergonomic graphical user interface. In an early CLARIN Deliverable (Hinrichs and Vogel, 2010), ANVIL (Kipp, 2014), CLAN (MacWhinney, 2000), ELAN (Sloetjes, 2014), EXMARaLDA (Schmidt and Wörner, 2014), FOLKER (Schmidt, 2012), Praat (Boersma, 2014), and Transcriber (Barras et al., 2000), have been identified as the annotation tools that are currently most relevant to the CLARIN community for this task. Most of them (CLAN and Praat being the exceptions) do work with XML based formats, but, so far, none of them “natively” operates on a TEI compliant format. In order to make the ISO/TEI standard work in practice, it is therefore essential to furnish users with an easy-to-use way of converting from a given tool format to the ISO/TEI format (ideally also the other way around, but this is more complex and will not be dealt with here[6]). The closer the tool format is to TEI’s general structure, the more straightforward this conversion will be.

Transcriber and FOLKER both use a data model which closely resembles the ISO/TEI approach insofar as it organizes transcripts as lists of speaker contributions (marked-up as <Turn> in Transcriber and as <contribution> in FOLKER) which can be seen as directly corresponding to TEI’s <u> elements. Whereas FOLKER (example 2a in Figure 6) allows additional markup for transcribed text underneath that unit (most importantly <w> for tokens), Transcriber represents the actual transcription text as plain

[5] See: http://www.parthenos-project.eu/ and http://www.parthenos-project.eu/Download/Deliverables/D4.1_Standardization_Survival_Kit.pdf

[6] The ISO/TEI format is potentially richer in information than any single of the tool formats taken into consideration. A conversion from ISO/TEI to a tool format therefore has to deal with all kinds of possible information loss.

character data and uses additional markup only for non-speech elements (1a). Both formats can be transferred to ISO/TEI (2b and 1b) as a more or less direct mapping of elements, without any fundamental structural changes. This can be achieved by simple XSLT transformations.

```
(1a) <Turn speaker="spk1" startTime="0.511" endTime="7.356">
        <Sync time="0.511"/><Event desc="souffle" type="noise" extent="instantaneous"/>
        <Sync time="1.593"/>hé bien euh bonsoir à tous et merci d'être restés si nombreux
        <Sync time="5.174"/>
        <Event desc="rire" type="noise" extent="instantaneous"/>
     </Turn>

(1b) <u who="#spk1" start="#T1" end="#T4">
         <vocal><desc>souffle</desc></vocal>
         <anchor synch="#T2"/>
         hé bien euh bonsoir à tous et merci d'être restés si nombreux
         <anchor synch="#T3"/>
         <vocal><desc>rire</desc></vocal>
     </u>

(2a) <contribution speaker-reference="IL" start-reference="TLI_33" end-reference="TLI_36">
        <w>zweihundert</w><w>ist</w><w>die</w><w>äh</w>
        <breathe type="in" length="1"/>
        <w>ist</w><w>die</w><w>planung</w>
        <time timepoint-reference="TLI_34"/>
        <pause duration="micro"/>
        <w>wenn</w><w>es</w><w>ausgebaut</w>
        <time timepoint-reference="TLI_35"/>
        <w>wird</w>
     </contribution>

(2b) <u who="#IL" start="#TLI_33" end="#TLI_36">
        <w xml:id="wd1e791">zweihundert</w>
        <w xml:id="wd1e799">ist</w>
        <w xml:id="wd1e801">die</w>
        <w xml:id="wd1e805">äh</w>
        <vocal>
           <desc>short breathe in</desc>
        </vocal>
        <w xml:id="wd1e808">ist</w>
        <w xml:id="wd1e810">die</w>
        <w xml:id="wd1e813">planung</w>
        <anchor synch="#TLI_34"/>
        <pause type="micro"/>
        <w xml:id="wd1e817">wenn</w>
        <w xml:id="wd1e819">es</w>
        <w xml:id="wd1e821">ausgebaut</w>
        <anchor synch="#TLI_35"/>
        <w xml:id="wd1e824">wird</w>
   </u>
```

Figure 6: Transformation of Transcriber (1) and FOLKER (2) file formats to ISO/TEI

CLAN's CHAT format works similarly in principle, with the important difference, however, that it is a plain text format and thus less directly usable as an input to an XSLT transformation. We currently use an existing CHAT-to-EXMARaLDA converter and transform its result to ISO/TEI.

The other formats differ from Transcriber, FOLKER and CHAT in that they are tier-based and thus do not provide a direct equivalent for the <u> element. For conversion of EXMARaLDA files to ISO/TEI, we rely on the concept of a "segment chain" – a maximally long sequence of contiguous annotations in a main tier – as the basic building block of a transcription. Segment chains are mapped to <u> elements and the respective contents of dependent tiers are then integrated in appropriate subordinate <spanGrp> elements. The principle is discussed in more depth in Schmidt (2005). Again, a single XSLT transformation is sufficient to achieve the conversion from EXMARaLDA (1a in Figure 8) to the ISO/TEI format (1b).

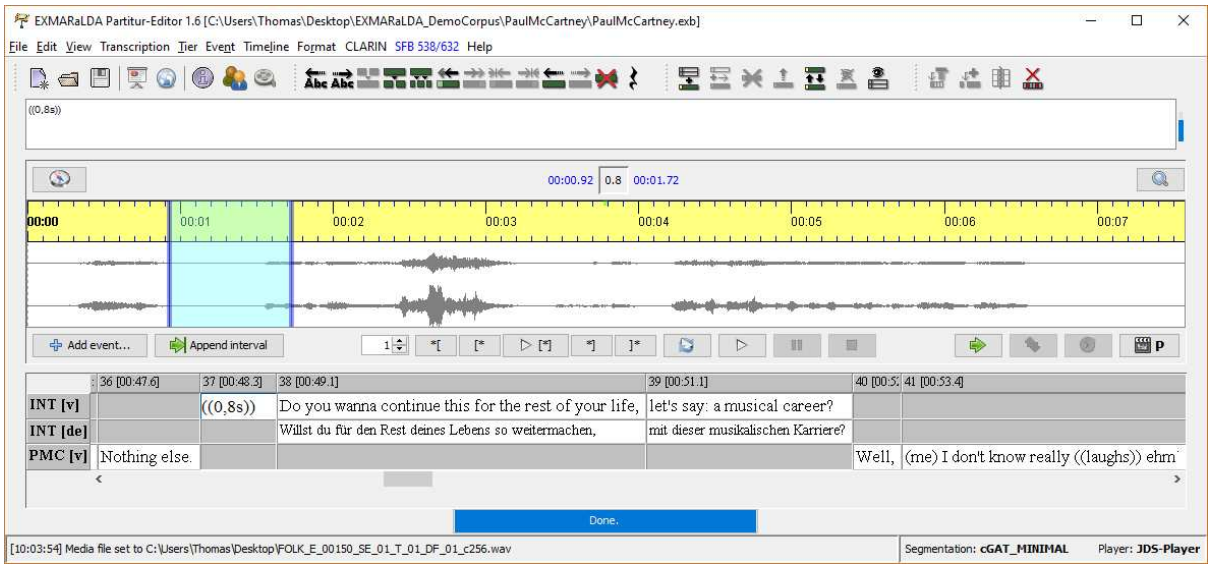

Figure 7: Musical score representation of EXMARaLDA's tier-based transcription format

```
(1a) <tier id="TIE0" speaker="SPK1" category="v" type="t" display-name="INT [v]">
        <event start="T37" end="T38">((0,8s))</event>
        <event start="T38" end="T39">Do you wanna continue this for the rest of your life, </event>
        <event start="T39" end="T40">let's say: a musical career? </event>
    </tier>
    <tier id="TIE2" speaker="SPK1" category="de" type="a" display-name="INT [de]">
        <event start="T38" end="T39">Willst du für den Rest deines Lebens so weitermachen, </event>
        <event start="T39" end="T40">mit dieser musikalischen Karriere? </event>
    </tier>
    <tier id="TIE3" speaker="SPK0" category="v" type="t" display-name="PMC [v]">
        <event start="T36" end="T37">Nothing else. </event>
        <event start="T40" end="T41">Well, </event>
        <event start="T41" end="T42">(me) I don't know really ((laughs)) ehm˙ </event>
    </tier>

(1b) <annotationBlock who="#SPK0" start="#T36" end="#T37">
         <u xml:id="u_d1e119">Nothing else. </u>
     </annotationBlock>
     <annotationBlock who="#SPK1" start="#T37" end="#T40">
         <u xml:id="u_d1e102">((0,8s))Do you wanna continue this for the rest of your life,
             <anchor synch="#T39"/>let's say: a musical career? </u>
         <spanGrp type="de">
             <span from="#T38" to="#T39">Willst du für den Rest deines
                   Lebens so weitermachen, </span>
             <span from="#T39" to="#T40">mit dieser musikalischen Karriere? </span>
         </spanGrp>
     </annotationBlock>
     <annotationBlock who="#SPK0" start="#T40" end="#T42">
         <u xml:id="u_d1e122">Well, (me) I don't know really ((laughs)) ehm˙ </u>
     </annotationBlock>
```

Figure 8: Transformation of EXMARaLDA file format to ISO/TEI

Praat's TextGrid format can be treated in a similar manner with, again, a detour via an existing Praat-to-EXMARaLDA converter because TextGrids are plain text, not XML, files. Since, however, Praat's data model only allows tiers to have a name (in the form of a simple string) and has no place for further structural information about tiers (such as speaker assignment or dependencies between transcription and annotation tiers), some information necessary for a TEI conversion will either have to be derived from ad-hoc tier naming conventions, or be added manually in a tool (such as EXMARaLDA) whose data model is sufficiently specific in this respect.

The situation is a bit more complex for ELAN's EAF format because of its more powerful possibilities of defining tier dependencies. In the context of the extended focus of the HZSK related to the associated CLARIN-D discipline-specific working group on Linguistic Fieldwork, Ethnology, and Language Typology, we are currently experimenting with existing EAF data sets from a language documentation

background, and can confirm that, at a minimum, an ELAN to ISO/TEI conversion should be possible as a rule on a per-corpus basis. We have not tackled ANVIL's file format yet.

Depending on the information available in the input format, some results of conversions from a tool format to ISO/TEI will have token information (example 2b in Figure 6) and some will not (examples 1b in Figures 6 and 8). Tokenization information – expressed in the ISO/TEI format most importantly through <w> markup – is, however, crucial for the majority of automatic language tools. Unfortunately, tokenization algorithms developed for written language are of limited use for spoken language transcriptions because they are not aware of the form and meaning of non-speech elements like pauses, descriptions of non-verbal behavior etc. Before a spoken language transcription is fed into a POS tagger or similar tools, a tokenization must therefore be carried out in order to separate (i.e. markup) "real" words from other, non-word elements. In the context of EXMARaLDA, we have developed a mechanism (called segmentation algorithm) which makes use of the regularities defined by transcription systems (such as "pauses are written as decimal numbers between round brackets") in order to achieve such a tokenization (again, see Schmidt, 2005 for further details). So far, EXMARaLDA is able to cleanly tokenize transcripts following the HIAT (Rehbein et al., 2004), GAT (Selting et al., 2009), cGAT (Schmidt et al., 2015), CHAT (MacWhinney, 2000) and DIDA (Klein & Schütte, 2001) transcription conventions. Ideally, these segmentation algorithms would have to be adapted to operate on the ISO/TEI (rather than the EXMARaLDA) format. They could then be offered (for instance, by a web service) as an additional processing step between the tool format conversion and further annotation through, say, a lemmatizer or POS tagger. We plan to address this requirement in the near future. For the time being, EXMARaLDA and its built-in segmentation algorithms for the different transcription systems can be used as an intermediary, for instance to transform a (non-tokenized) Transcriber file via EXMARaLDA, where a HIAT segmentation algorithm is applied, to a tokenized ISO/TEI file.

Export filters from the tool to the ISO/TEI format have been built directly into EXMARaLDA and FOLKER. For other tool formats or batch conversion of entire corpora, the EXMARaLDA distribution provides TEI-Drop (see Figure 9), a droplet desktop application onto which users can drag and drop a number of input files (in Transcriber, FOLKER, EXMARaLDA, CHAT or EAF), specify a few parameters such as which segmentation algorithm to use and where to write the output, and which will then perform the conversions in a single step.

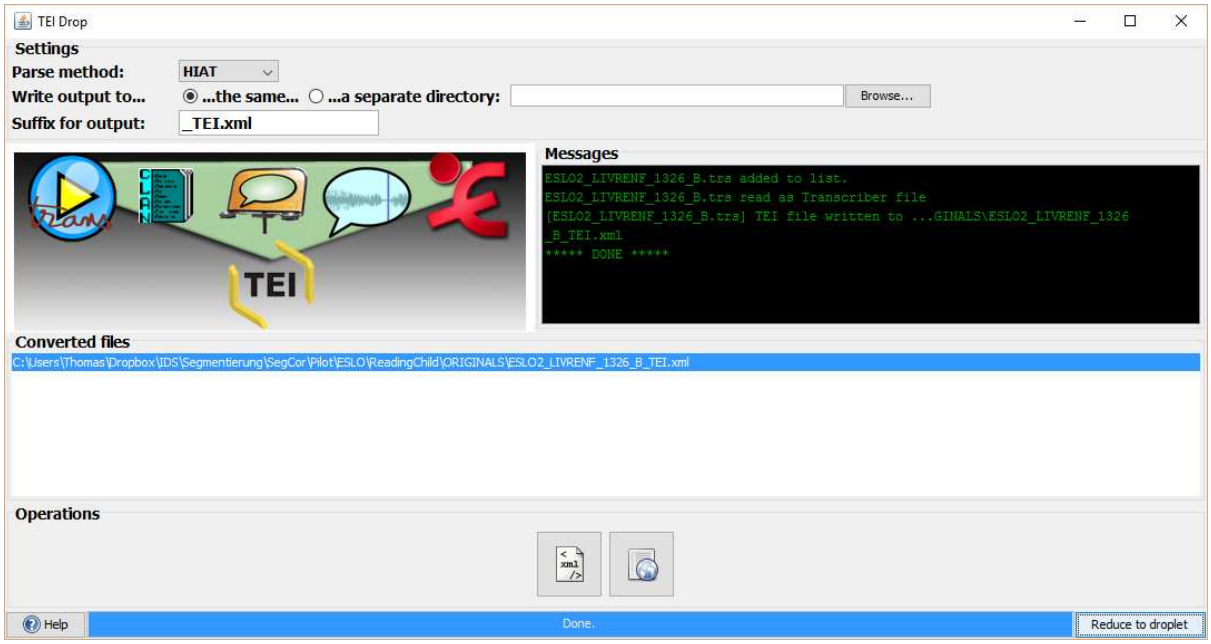

Figure 9: Screenshot of TEI Drop

In order for the standard to be adopted on a larger scale, it is crucial not only that the possibility of converting existing data is provided, but also that essential existing language resources are made available in that format to the community. The Archive for Spoken German with its DGD platform[7] (Schmidt,

[7] See: http://dgd.ids-mannheim.de

2014) as the central provider of German oral corpora, as well as the Hamburg Centre for Language Corpora with its repository of multilingual spoken data[8] (Jettka and Stein, 2014) are two certified CLARIN centers that are in principle ready to offer their resources (more than 5000 hours of material, altogether) in the ISO/TEI format.

Figure 10: Export options for transcript excerpts in the Database for Spoken German (DGD)

For the time being, that format will be an additional option beside the established formats favored by the respective centers (as in Figure 10). In the mid-term, however, we aim at making ISO/TEI the central format for disseminating our resources. Ongoing or recently piloted cooperations with international partners – most importantly the ESLO corpus in Orléans (Eshkol-Taravella et al., 2012), the CLAPI database in Lyon (Groupe ICOR, 2017), the Speech Island Portal in Austin/Texas[9], and the spoken components of the Australian National Corpus[10] – may, we hope, lead to a wider distribution of the standard on an international level.

Another crucial aspect for the acceptance of the standard is the interoperability with existing tools to further annotate, to query, and to analyze transcribed data. Support for ISO/TEI as an import/export format in WebAnno[11] (Eckart de Castilho et. al, 2014) is planned to be developed starting this spring, which would enable web-based manual annotation using annotation layers such as dependency parsing and co-reference resolution. For the search and visualization tool ANNIS (Krause and Zeldes, 2016)[12], a prototype Pepper (Zipser and , 2010) conversion module has been implemented. Finally, the Solr-based Multi Tier Annotation Search (MTAS, Brouwer and Kemps-Snijders, 2016)[13] system developed at the Meertens Instituut has very recently been extended to also index the ISO/TEI format and a version including ISO/TEI transcription samples is being tested. Support for the ISO/TEI format would enable a powerful CQL-based search to be used with the various transcription formats for which conversion methods exist.

[8] See: http://hdl.handle.net/11022/HZSK-0000-0000-2C76-B-REPOSITORY.
[9] See: http://speechislands.org
[10] See, for instance: https://www.ausnc.org.au/corpora/gcsause
[11] See: https://webanno.github.io/
[12] See: http://corpus-tools.org/annis/
[13] See: https://meertensinstituut.github.io/mtas/

4 Using TCF-based Web Services in WebLicht

Currently, all services integrated in WebLicht operate on version 0.4 of the Text Corpus Format (TCF)[14]. In order for an ISO/TEI document to be sent through a WebLicht tool chain, it therefore needs to be encoded in TCF before the first step of the chain is applied, and the result of the chain needs to be decoded from TCF to ISO/TEI after the last step of the chain. TCF is a format based on the "stream of tokens" idea and assumes that the basic structure of any document is a linear sequence of token elements:

> The tokens layer is [thus] the main anchor layer among TextCorpus layers, i.e. all other layers (with the exception of the text layer) directly or indirectly (via other layers) reference tokens by referencing token identifiers. [Introduction to section 3 of the TCF specification]

TCF thus does not have in its basic structure any means of representing information that is crucial to spoken language transcription, such as time alignment and speaker assignment, and it also does not provide the possibility to distinguish different types of tokens (such as words vs. non-speech descriptions) on its basic layer. ISO/TEI-to-TCF conversion is therefore bound to be lossy – not all information contained in a transcription can be meaningfully mapped onto some TCF element.

Our general approach for the TCF encoding of ISO/TEI files is therefore:

1. To only map those elements of an ISO/TEI file which have a more or less direct equivalent in TCF. Basically, this boils down to mapping <w> elements (and, as the case may be <pc> elements containing punctuation) in ISO/TEI to <token> elements in TCF.[15]
2. To assume that most services in WebLicht will work fine with this reduced set of information, i.e. will produce useful results, although some types of information have been removed.
3. To keep on the <tokens> layer the @xml:id attributes of the original document (in an @ID attribute of the individual <token> elements).
4. To keep in the <textSource> element the entire original ISO/TEI document.

When such a TCF document is fed through a tool chain in WebLicht, any resulting additional annotation layers can then be remapped to a suitable ISO/TEI form via the memorized @xml:id attributes. Since the original document in the <textSource> element is also fed through the tool chain unchanged, this decoding step can be stateless as required by the SOA architecture of CLARIN in general.

Many tools in WebLicht require information about sentence boundaries. Since spoken language transcriptions, as a rule, do not operate with the concept of a "sentence", a further important question in the encoding process is therefore which elements of the source document can be meaningfully mapped onto the <sentences> layer in TCF. Our approach to this question is very straightforward: if a segmentation of the transcript in the form of <seg> elements directly underneath the <u> elements exists, we use that segmentation as a sentence equivalent. Although, semantically, these entities (such as intonation phrases or speech acts) are usually not sentences in the written-language sense of the word, they are usually sufficiently similar to sentences in order to function as an adequate entity in their stead. In the absence of <seg> elements, we use the entire <u> element as a sentence equivalent.

Figure 11 illustrates how an excerpt of an ISO/TEI transcription (1) is encoded in TCF (2), supplemented with the result of a POS tagger (3) and then redecoded to ISO/TEI (4).

[14] See: http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format

[15] Alternatively, one might map onto TCF's <token> element not the character data of <w> elements, but the content of elements in <spanGrp> which represent orthographic normalised forms (see figure 5 above), if such an annotation level exists. This has a potential of improving annotation results in WebLicht, because individual tools in the chain would then operate on the standard orthography they usually expect. This could be implemented as a parameter passed to the TEI to TCF converter or even be decided autonomously by the web service. We currently refrain from such an implementation because our focus here is on establishing a general mechanism for CLARIN integration of spoken language data.

```
(1) <u who="MJ" start="#T0" end="#T2">
   <seg type="intonation-phrase" subtype="falling">
        <w xml:id="w1">I</w>
        <vocal><desc>cough</desc></vocal>
        <w xml:id="w2">see</w>
        <w xml:id="w3">a</w>
        <w xml:id="w4">door/w>
    </seg>
    <anchor synch="#T1"/>
    <seg type="intonation-phrase" subtype="falling">
        <w xml:id="w5">I</w>
        <pause dur="PT0.3S"/>
        <w xml:id="w6">want</w>
        <w xml:id="w7">to</w>
        <w xml:id="w8">paint/w>
        <w xml:id="w9">it</w>
        <unclear>
            <choice>
                <w xml:id="w10">black</w>
                <w xml:id="w11">blue</w>
            </choice>
        </unclear>
    </seg>
</u>

(2) <TextCorpus>
    <text>I see a door I want to paint it black blue</text>
    <tokens>
       <token ID="w1">I</token>
       <token ID="w2">see</token>
       <token ID="w3">a</token>
       <token ID="w4">door</token>
       <token ID="w5">I</token>
       <!-- [...] -->
       <token ID="w9">it</token>
       <token ID="w10">black</token>
       <token ID="w11">blue</token>
    </tokens>
    <sentences>
       <sentence ID="s_1" tokenIDs="w1 w2 w3 w4"/>
       <sentence ID="s_2" tokenIDs="w5 w6 w7 w8 w9 w10 w11"/>
    </sentences>
    <textSource type="application/tei+xml;format-variant=tei-iso-spoken;tokenized=1">
        <![CDATA[
            <TEI xmlns="http://www.tei-c.org/ns/1.0">
               [...]
               <u who="MJ" start="#T0" end="#T2">
               [...]
            </TEI>
        ]]>
    </textSource>
<TextCorpus>

(3) <TextCorpus>
    <!-- [...] -->
    <POStags tagset="stts">
            <tag ID="pt_0" tokenIDs="w1">PPER</tag>
            <tag ID="pt_1" tokenIDs="w2">V</tag>
            <tag ID="pt_2" tokenIDs="w3">DET</tag>
            <tag ID="pt_3" tokenIDs="w4">NN</tag>
            <!-- [...] -->
            <tag ID="pt_11" tokenIDs="w11">ADJ</tag>
        </POStags>
    <!-- [...] -->
</TextCorpus>

(4) <annotationBlock who="MJ" start="#T0" end="#T2" xml:id="ab1">
   <u xml:id="u1">
        <seg type="intonation-phrase" subtype="falling" xml:id="seg1">
                <w xml:id="w1">I</w>
                <vocal xml:id="voc1"><desc>cough</desc></vocal>
```

```
                <w xml:id="w2">see/w>
                <w xml:id="w3">a/w>
                <w xml:id="w4">door/w>
        </seg>
    </u>
    <spanGrp type="pos">
        <span from="#w1" to="#w1">PPER</span>
        <span from="#w2" to="#w2">V</span>
        <span from="#w3" to="#w3">DET</span>
        <span from="#w4" to="#w4">NN</span>
    </spanGrp>
</annotationBlock>
```

Figure 11: Encoding, WebLicht processing and decoding of an ISO/TEI file

5 CLARIN integration

An important practical question concerning the integration of the conversion services into the WebLicht context (and, ultimately, into the CLARIN infrastructure in general) is how to inform other services of the format of the admissible input (what the service "consumes") and the expected output (what the service "produces") of any given single converter. MIME type[16] in CMDI specifications had been used for that purpose before, but without a clear, commonly agreed guideline, resulting in several unsolved challenges:

1. MIME types have to be sufficiently specific for the application that is meant to handle the respective file format. For example, a MIME type like "application/tei+xml", although perfectly valid, will not be sufficient for the converters described here, because those expect a specific type of TEI file (i.e. one that conforms to the ISO standard) and will fail when confronted with other TEI variants.
2. Only one MIME type should be used consistently per given file format. For instance, it would be possible to describe an ELAN file either as "text/x-eaf+xml" (to be found in several existing CLARIN records) or as "application/xml" (more in line with current recommendations for XML files). A web service must at least know all possible variants or, better still, be sure that exactly one variant is used in the infrastructure.
3. The use of registered MIME types should be preferred over non-registered MIME types. Non-registered MIME types (e.g. using the "x-" prefix) have been used ad hoc to describe specific annotation formats, but using non-registered MIME types in this context is problematic, since they were "intended exclusively for use in private, local environments"[17].

After discussions involving the CLARIN task force on WebLicht/TEI, the CLARIN-D developer group and the CLARIN standards committee, we settled for the MIME type assignments in Table 1. These MIME types comply with current recommendations and best practices[18] . The first part always consists of a registered MIME type so that even applications unaware of the specific linguistic annotation formats still have a chance to do some useful processing of the data (e.g. a web browser will recognize an XML file as such and display its DOM tree). In the second part, an additional parameter "format-variant" is used to provide the more detailed format information needed by the web services described here.

[16] For reasons of consistency, we use "MIME type" to refer to media types (as they are now officially called) throughout this paper.
[17] See: https://tools.ietf.org/html/rfc6838#section-3.4
[18] See: https://tools.ietf.org/html/rfc2046 and https://www.w3.org/TR/webarch/#xml-media-types

Format	MIME type
Text Corpus Format (*.tcf)	text/tcf+xml or application/xml; format-variant=weblicht-tcf[19]
ISO/TEI for transcriptions of spoken language	application/tei+xml; format-variant=tei-iso-spoken[20]
EXMARaLDA Basic transcription (*.exb)	application/xml; format-variant=exmaralda-exb
Transcriber annotation file (*.trs)	application/xml; format-variant=transcriber-trs
FOLKER transcription (*.flk / *.fln)	application/xml; format-variant=folker-fln
CHAT transcription file (*.cha)	text/plain; format-variant=clan-cha[21]
ELAN Annotation File (*.eaf)	application/xml; format-variant=elan-eaf[22]
Praat TextGrid (*.textGrid)	text/plain; format-variant=praat-textgrid[23]

Table 1: MIME types for different annotation formats

6 Using WebLicht tool chains in practice

The different converters are made available as individual web services, hosted by the CLARIN cener HZSK and exposed via a CMDI description (e.g. http://hdl.handle.net/11022/0000-0000-9ABA-1 for a development version of the EXMARaLDA-to-ISO/TEI converter) to CLARIN services such as the Virtual Language Observatory and WebLicht. Once fully integrated into CLARIN in that way[24], WebLicht will recognize the different annotation tool formats and offer users the ISO/TEI conversion service as a possible first conversion step. By applying the ISO/TEI-to-TCF converter to the result, the rest of the annotation services in WebLicht become available for the data. In a final step, the result of an annotation chain can be decoded to ISO/TEI again.

The WebLicht web interface is a good way of getting acquainted with and testing different annotation tool chains for a given piece of data. However, for many users who have already decided on a tool chain, it may be more convenient to apply it directly inside the tool with which the data are created. For that purpose, we have integrated functionality for calling the SaaS variant of WebLicht – WebLicht as a Service (WaaS[25]) – out of the EXMARaLDA Partitur-Editor desktop tool. A typical workflow that is served by that functionality would proceed as follows:

1. A transcription is created in the EXMARaLDA Partitur-Editor, or imported there from a CLAN, ELAN, FOLKER, Praat or Transcriber file.
2. A tool chain for WebLicht is constructed by:
   a. Exporting a TCF version of the transcription from the Partitur-Editor
   b. Uploading that TCF version to the WebLicht web interface
   c. Constructing and testing a suitable tool chain inside the web interface
   d. Saving the tool chain locally
3. Users obtain a valid WaaS key[26] using Single sign-on (based on Shibboleth/CLARIN SPF)

[19] The latter variant is consistent with the patterns for the other MIME types. For reasons of backward compatibility, however, the WebLicht developers prefer to stick to the first variant

[20] Note that "application/tei+xml" is a registered MIME type, and that the MIME type for the TEI format defined at the Deutsches Textarchiv (DTA) for encoding (historical) written text has been decided upon in this process to be "application/tei+xml; format-variant=tei-dta"

[21] An additional parameter "charset" might be used to specify the encoding, as is common for plain text files

[22] This is a suggestion only - none of the services described here consumes or produces eaf files, so the MIME type is nowhere used

[23] This is a suggestion only - none of the services described here consumes or produces textGrid files, so the MIME type is nowhere used. An additional parameter "charset" might be used to specify the encoding, as is common for plain text files

[24] All conversion services described here are now online in a beta version (see Appendix). Full WebLicht integration is still pending, but we expect it to be complete by the time this paper is due to be published.

[25] See: https://weblicht.sfs.uni-tuebingen.de/WaaS/

[26] See: https://weblicht.sfs.uni-tuebingen.de/WaaS/apikey

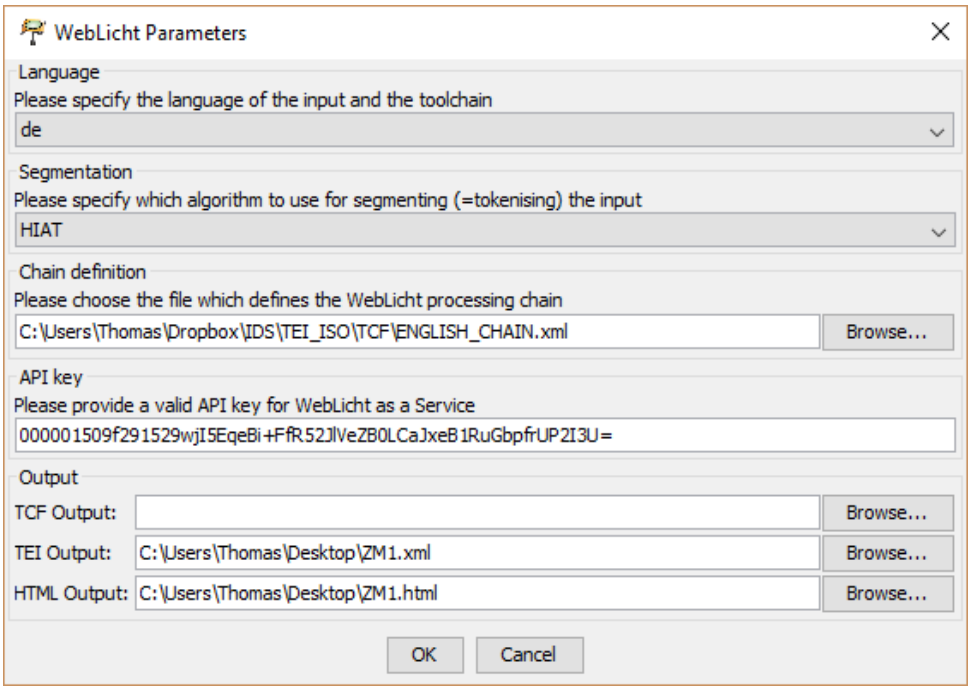

Figure 12: Parameter Dialog for WebLicht in the EXMARaLDA Partitur-Editor

The tool chain can then be used on the original data and further datasets from the same collection by calling "WebLicht…" from the CLARIN menu in the Partitur-Editor. This will bring up the dialog from Figure 12, in which parameters for WaaS can be specified.

After clicking on "OK", EXMARaLDA will perform the necessary conversions to TCF, send the resulting file to WaaS alongside the specified parameters, receive the reply from the service and store it locally, either as the TCF obtained from WaaS, as an ISO/TEI file (after applying the decoding converter), or as an HTML file representing visually the different annotations added in WebLicht.

7 ISO/TEI-based Web Services

While TCF encoding and decoding is a workable solution for using existing services in WebLicht on spoken language data, it is, in the long run, not an optimal way of dealing with such data.

Obviously, the information lost in the TCF encoding process has a potential value for many automatic processing methods. For example, a POS tagger optimized for spoken language might make use of n-gram statistics involving information about pauses, or a lemmatizer may ignore, or treat in a special manner, defect tokens such as aborted words. Likewise, some services may want to use orthographic normalizations (see figure 5 above), time alignment information or even the corresponding section of the audio or video signal data. These pieces of information are available all in the ISO/TEI format, but not in TCF.

Moreover, in some cases, basic assumptions tacitly included in TCF's data model may turn out to be overly simplified when applied to spoken language data. For instance, TCF specifies "the language of the data [emphasis added]"[27]inside a single attribute @lang on the <textCorpus> element. Multilingual interactions, such as interpreted talk, however, contain by definition data in at least two languages, and there is no way of telling a TCF based tool which part of the document is in which language. In the ISO/TEI format, by contrast, an @xml:lang can optionally be used on every element in the document to provide such information.

Ideally, automatic annotation tools and services optimized for spoken language transcription would therefore operate directly on the ISO/TEI format, and the detour via TCF conversion would then become unnecessary. One existing example for this is an adaptation of TreeTagger and the STTS tagset for use with transcriptions from the FOLK project (see Westpfahl and Schmidt, 2016). We are currently working on turning this mechanism into a CLARIN compliant (but not TCF compatible) web service. Similar tools developed in the context of the AGD and HZSK corpora, such as a tool for automatic orthographic

[27] See: http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format

normalization of transcripts, and a tool for fine-aligning transcripts via MAUS, could be treated in an analogous manner.

8 Conclusion

As the present contribution shows, there are ways of making existing CLARIN components and architectural concepts that were originally or mainly developed for canonical written language data usable also for spoken language data. The issue of standardization is crucial in this task, since, without a widely used and sufficiently specified common basis, the number of processing steps needed to convert a given piece of data from and to a form usable by WebLicht would multiply. The newly published ISO/TEI standard provides such a common basis for various established tool formats and transcription conventions for spoken language data. While being able to represent common time-based annotation formats consistently and without information loss, its hierarchical structure facilitates conversion to traditional formats for canonical written language data.

Since conversion to these formats is however not possible without information loss, we believe that spoken language data should ideally be processed by services aware of and capable of interpreting the complexity and peculiarities inherent to this type of data. In a wider scope, CLARIN as an infrastructure also needs to consider and meet the requirements and expectations of various research communities working with different types of non-canonical and/or non-written language data. Though this is a challenging task to take on, considering the increasing availability of audio and video data and thus its increasing importance in linguistic and language based research, it seems crucial to keep up with this development in order for CLARIN to become and remain equally available and important to researchers from all targeted fields.

Another critical aspect of a successful digital research infrastructure for language resources and technology is a common set of relevant standards recognized by all components of the infrastructure. This requires relevant file formats to be identified and described on a level specific enough to allow for automatic processing. Such descriptions are necessary to be able to inform not only other components within the infrastructure – but also the users – about the technical characteristics of language resources and the suitability and interoperability of existing tools and services. The experiences with processing various flavors of TEI within WebLicht may serve as a useful pilot experiment for future efforts to provide researchers with reliable guidance in choosing tools and services for their workflows and processing pipelines.

Consequently, there is also work to be done regarding the various widely used de facto standard annotation formats in order to allow for a consistent use of MIME types and possible complementary descriptions within the CLARIN infrastructure. As an international organization with a strong network, CLARIN could play an important role in this matter, creating an impact that goes beyond the infrastructure itself, e.g. through coordination of the registration of relevant formats with existing standardization bodies.

References

[Bański et al. 2016] Bański, Piotr, Bertrand Gaiffe, Patrice Lopez, Simon Meoni, Laurent Romary, Thomas Schmidt, Peter Stadler, and Andreas Witt. 2016. In Claudia Resch, Vanessa Hannesschläger, and Tanja Wissik, editors, *TEI Conference and Members' Meeting 2016 – Book of Abstracts*, Vienna, pages 35-37. http://tei2016.acdh.oeaw.ac.at/sites/default/files/TEIconf2016_BookOfAbstracts.pdf

[Barras et al. 2000] Barras, Claude, Edouard Geoffrois, Zhibiao Wu, and Mark Liberman. 2001. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communicatin – Special issue on Speech Annotation and Corpus Tools*, 33(1-2):5–22. http://languagelog.ldc.upenn.edu/myl/Barras2001.pdf. DOI: 10.1016/S0167-6393(00)00067-4.

[Bird and Liberman 2001] Bird, Steven and Mark Liberman. 2001. A formal framework for linguistic annotation. *Speech Communicatin – Special issue on Speech Annotation and Corpus Tools*, 33(1-2):23–60. http://languagelog.ldc.upenn.edu/myl/BirdLiberman2001.pdf. DOI: 10.1016/S0167-6393(00)00068-6.

[Boersma 2014] Boersma, Paul. 2014. The Use of Praat in Corpus Research. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *The Oxford Handbook of Corpus Phonology*. Oxford University Press, Oxford, UK. DOI: 10.1093/oxfordhb/9780199571932.013.016.

[Brouwer and Kemps-Snijders 2016] Brouwer, Maththijs and Kemps-Snijders, Marc. 2016. A SOLR/Lucene based Multi Tier Annotation Search solution. *Proceedings of the CLARIN Annual Conference* (CAC), 2016, Aix, France. https://www.clarin.eu/sites/default/files/brouwer-kempssnijdersCLARIN2016_paper_21.pdf

[Eckart de Castilho et. al 2014] Eckart de Castilho, Richard, Biemann, Chris, Gurevych, Irina and Yimam, S.M. 2014. WebAnno: a flexible, web-based annotation tool for CLARIN. In *Proceedings of the CLARIN Annual Conference* (CAC) 2014, Soesterberg, Netherlands.

[Eshkol-Taravella et. al 2012] Eshkol-Taravella, Iris, Oliver Baude, Denis Maurel, Linda Hriba, Céline Dugua, and Isabelle Tellier. 2012. Un grand corpus oral «disponible»: le corpus d'Orléans 1968-2012. *Ressources linguistiques libres*, 52(3/2011):17–46. https://www.atala.org/IMG/pdf/Eshkol-TAL52-3.pdf.

[Groupe ICOR 2017] GROUPE ICOR (Heike Baldauf-Quilliatre, Isabel Colón de Carvajal, Carole Etienne, Emilie Jouin-Chardon, Sandra Teston-Bonnard, and Véronique Traverso) (in press). CLAPI, une base de données multimodale pour la parole en interaction : apports et dilemmes. In Mathieu Avanzi, Marie-José Béguelin, and Federica Diémoz, editors, *Corpus de français parlés et français parlés des corpus,* Cahiers Corpus, Université de Neuchâtel.

[Hinrichs and Vogel 2010] Hinrichs, Erhard and Iris Vogel. 2010. CLARIN - Interoperability and Standards. In *CLARIN devliverable D5.C-3*. http://www-sk.let.uu.nl/u/D5C-3.pdf.

[Hinrichs et al. 2010] Hinrichs, Marie, Thomas Zastrow, and Erhard Hinrichs. 2010. WebLicht: Web-based LRT Services in a Distributed eScience Infrastructure. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valetta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/270_Paper.pdf.

[IETF] Internet Engineering Task Force. 2013. *Media Type Specifications and Registration Procedures – Best Practices*. https://tools.ietf.org/html/rfc6838#section-3.4.

[ISO 2016] ISO 2462:2016. *Language resource management – Transcription of spoken language*. http://www.iso.org/iso/catalogue_detail.htm?csnumber=37338.

[Jettka & Stein 2014] Jettka, Daniel, Daniel Stein. 2014. The HZSK Repository: Implementation, Features, and Use Cases of a Repository for Spoken Language Corpora. *D-Lib Magazine*, 20(9/10). http://www.dlib.org/dlib/september14/jettka/09jettka.html. DOI: 10.1045/september2014-jettka.

[Kipp 2014] Kipp, Michael. 2014. ANVIL: A Universal Video Research Tool. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *The Oxford Handbook of Corpus Phonology*. Oxford University Press, Oxford, UK. DOI: 10.1093/oxfordhb/9780199571932.013.024.

[Kisler et al. 2010] Kisler, Thomas, Florian Schiel, and Han Sloetjes. 2012. Signal processing via web services: the use case WebMAUS. In *Proceedings of Digital Humanities 2012*, Hamburg, pages 30–34. http://clarin-d.de/images/workshops/proceedingssoasforthehumanities.pdf.

[Klein and Schütte 2001] Klein, Wolfgang, Wilfried Schütte. 2001. *Transkriptionsrichtlinien für die Eingabe in DIDA*. Institut für Deutsche Sprache, Mannheim, Germany. http://agd.ids-mannheim.de/download/dida-trl.pdf.

[Krause and Zeldes 2016] Krause, Thomas and Zeldes, Amir. 2016. ANNIS3: A new architecture for generic corpus query and visualization. In *Digital Scholarship in the Humanities 2016 (31).* http://dsh.oxfordjournals.org/content/31/1/118.

[MacWhinney 2000] MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Psychology Press, New York, USA.

[Menke et al. 2015] Menke, Peter, Farina Freigang, Thomas Kronenberg, Sören Klett, and Kirsten Bergmann. 2015. First Steps towards a Tool Chain for Automatic Processing of Multimodal Corpora. *Journal of Multimodal Communication Studies*, 2:30–43. http://jmcs.home.amu.edu.pl/wp-content/uploads/2015/09/Menke_et_al_2014_JMCS.pdf.

[Parisse and Morgenstern 2010] Parisse, Christophe and Aliyah Morgenstern. 2010. A multi-software integration platform and support for multimedia transcripts of language. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* [Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality], Valetta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/.

[Rehbein et al. 2004] Rehbein, Jochen, Thomas Schmidt, Bernd Meyer, Franziska Watzke, and Annette Herkenrath. 2004. Handbuch für das computergestützte Transkribieren nach HIAT. In *Arbeiten zur Mehrsprachigkeit*, volume 56. http://www.exmaralda.org/files/azm_56.pdf.

[Schmidt 2005] Schmidt, Thomas. 2005. Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech. In *Arbeiten zur Mehrsprachigkeit (Folge B)*, volume 62. http://www.exmaralda.org/files/SFB_AzM62.pdf.

[Schmidt et al. 2009] Schmidt, Thomas, Susan Duncan, Oliver Ehmer, Jeffrey Hoyt, Michael Kipp, Dan Loehr, Magnus Magnusson, Travis Rose, and Han Sloetjes. 2009. An Exchange Format for Multimodal Annotations. In Michael Kipp, Jean-Claude Martin, Patrizia Paggio, Dirk Heylen, editors, *Multimodal Corpora*. Springer, Berlin, Germany. http://link.springer.com/chapter/10.1007/978-3-642-04793-0_13. DOI: 10.1007/978-3-642-04793-0_13.

[Schmidt 2011] Schmidt, Thomas. 2011. A TEI-based Approach to Standardising Spoken Language Transcription. *Journal of the Text Encoding Initiative*, 1:1–22. http://jtei.revues.org/142. DOI: 10.4000/jtei.142.

[Schmidt 2012] Schmidt, Thomas. 2012. EXMARaLDA and the FOLK tools. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. http://www.lrec-conf.org/proceedings/lrec2012/pdf/529_Paper.pdf.

[Schmidt 2014] Schmidt, Thomas. 2014. The Database for Spoken German – DGD2. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. http://www.lrec-conf.org/proceedings/lrec2014/pdf/171_Paper.pdf.

[Schmidt and Wörner 2014] Schmidt, Thomas and Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *The Oxford Handbook of Corpus Phonology*. Oxford University Press, Oxford, UK. DOI: http://dx.doi.org/10.1093/oxfordhb/9780199571932.013.030.

[Schmidt et al. 2015] Schmidt, Thomas, Wilfried Schütte, and Jenny Winterscheid. 2015. *cGAT. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2)*. https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/4616.

[Selting et al. 2009] Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 10:353–402. http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf.

[Sloetjes 2014] Sloetjes, Han. 2014. ELAN: Multimedia Annotation Application. In Jacques Durand, Ulrike Gut, and Gjert Kristoffersen, editors, *The Oxford Handbook of Corpus Phonology*. Oxford University Press, Oxford, UK. DOI: 10.1093/oxfordhb/9780199571932.013.019.

[TEI] Text Encoding Initiative. 2015. *Guidelines*. http://www.tei-c.org/Guidelines/.

[Westpfahl and Schmidt 2016] Westpfahl, Swantje and Thomas Schmidt. 2016. FOLK-Gold – A GOLD standard for Part-of-Speech-Tagging of Spoken German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia. http://www.lrec-conf.org/proceedings/lrec2016/pdf/397_Paper.pdf.

[Zipser and Romary 2010] Zipser, Florian and Romary, Laurent. 2010. A model oriented approach to the mapping of annotation formats using standards. In: Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010. Malta. http://hal.archives-ouvertes.fr/inria-00527799/en/.

Appendix A. Details for the web services

Beta versions of the conversion services are available as listed below, the source code for all services is available on GitHub: https://github.com/hzsk/HZSK-CLARIN-Services/.

EXMARaLDA to ISO/TEI conversion

PID:	http://hdl.handle.net/11022/0000-0000-9A6C-A
Consumes:	application/xml; format-variant=exmaralda-exb
Produces:	application/tei+xml; format-variant=tei-iso-spoken
Parameters:	seg - the segmentation algorithm to be used, one of (HIAT|cGAT|CHAT)
	lang - the language of the document, two letter ISO language code

FOLKER to ISO/TEI conversion

PID:	http://hdl.handle.net/11022/0000-0001-B538-4
Consumes:	application/xml; format-variant=folker-fln
Produces:	application/tei+xml; format-variant=tei-iso-spoken
Parameters:	lang - the language of the document, two letter ISO language code

Transcriber to ISO/TEI conversion

PID:	http://hdl.handle.net/11022/0000-0001-B539-3
Consumes:	application/xml; format-variant=transcriber-trs
Produces:	application/tei+xml; format-variant=tei-iso-spoken
Parameters:	seg - the segmentation algorithm to be used, one of (HIAT|cGAT|CHAT)
	lang - the language of the document, two letter ISO language code

CHAT to ISO/TEI conversion

PID:	http://hdl.handle.net/11022/0000-0001-B53A-2
Consumes:	text/plain;format-variant=clan-cha
Produces:	application/tei+xml; format-variant=tei-iso-spoken
Parameters:	seg - the segmentation algorithm to be used, one of (HIAT|cGAT|CHAT)
	lang - the language of the document, two letter ISO language code

ISO/TEI to TCF conversion

PID:	http://hdl.handle.net/11022/0000-0001-B53B-1
Consumes:	application/tei+xml; format-variant=tei-iso-spoken
Produces:	application/xml; format-variant=weblicht-tcf

TCF to ISO/TEI conversion

PID:	http://hdl.handle.net/11022/0000-0001-B53C-0
Consumes:	application/xml; format-variant=weblicht-tcf
Produces:	application/tei+xml; format-variant=tei-iso-spoken