

IIIF ビューワ「Mirador 動画アノテーション対応版」の改良： 動画アノテーションツール ELAN との連携を目指して

高橋 洋成¹ 永崎 研宣² 本間 淳³

概要：動画アノテーションツール ELAN で作成された「注釈付き動画」を言語教育の教材として活用するため、本研究は IIIF ビューワ「Mirador 動画アノテーション対応版」を改良し、注釈テキストの自動スクロール機能や字幕表示機能を追加した。また、ELAN の注釈データを IIIF 記述へ変換するにあたり、副産物として様々な形式のデータを得ることができ、多様なプラットフォームを見据えた注釈テキストの「エコシステム」構築の可能性が示された。

Improvements to the IIIF viewer ‘Mirador Video Annotation Support Version’: Toward Integration with ELAN, a Video Annotation Tool

Abstract: In order to utilize ‘annotated videos’ created with ELAN (a video annotation tool) as teaching materials for language education, this study improved the IIIF viewer ‘Mirador Video Annotation Support Version’ and added functions for automatically scrolling annotated text and displaying subtitles. In addition, in converting ELAN annotation format to IIIF resources, we were able to obtain data in various formats as a byproduct, indicating the possibility of building an ‘ecosystem’ of annotation texts for various platforms.

1. はじめに

1.1 背景

Ethnologue (第 25 版) によれば、エチオピア連邦民主共和国では話し言葉として 91 言語が使用されている [1]。これらのうち、エチオピア憲法は「アムハラ語を作業語とする」と定めるのみで公用語を規定してはいない。また、11 ある州の中にはアムハラ語に加えて独自の作業語を定める所もある。

このように多言語国家であるエチオピアでは、近年アムハラ語教育だけでなく母語教育（「母語の教育」「母語による教育」の両方）を推進する風潮が高まっている。しかし、91 ある言語の半数以上は正書法の確立はおろか、十分な調査研究も未だなされていないのが現状である。

筆者（高橋）が属している JAEL (Japan Association of Ethiopian Languages) は、エチオピア少数言語の調査研究

を主とする研究者が集まっている。2018 年度より、JAEL はアディス・アベバ大学の言語研究者らと連携し、少数言語を母語とする人々のための教材開発プロジェクトを開始した。これまで JAEL が蓄積してきた言語資料は話者の音声、映像、動画、その音訳、翻訳、語釈、語彙集、文法記述など多岐にわたる [2]。今後これらをデジタル・アーカイブ化し、幅広い利活用のための基盤構築を目指していくが、教材作成はその成果の 1 つになる。

1.2 ELAN

母語教育教材の 1 つとして、動画に対して音訳、翻訳、語釈等の字幕を付けた「字幕付き動画」の作成を進めている。具体的には ELAN (EUDICO Linguistic Annotator) によって動画に音訳、翻訳等の注釈テキストを付与する作業を行っている [3]。図 1 に示すように、ELAN は音声の波形データを目視しながら複数層の注釈テキストを付与できる点にその特徴がある。音訳、翻訳等と同じ時間帯に重ねて付与することができ、さらに複数の話者に対してそれを行うこともできる。したがってモノログにもダイアログにも対応可能である。

ただし、ELAN には「字幕付き動画」そのものを作成す

¹ 東京外国語大学アジア・アフリカ言語文化研究所
The Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies

² 人文情報学研究所
International Institute for Digital Humanities

³ フェリックス・スタイル
FLX Style

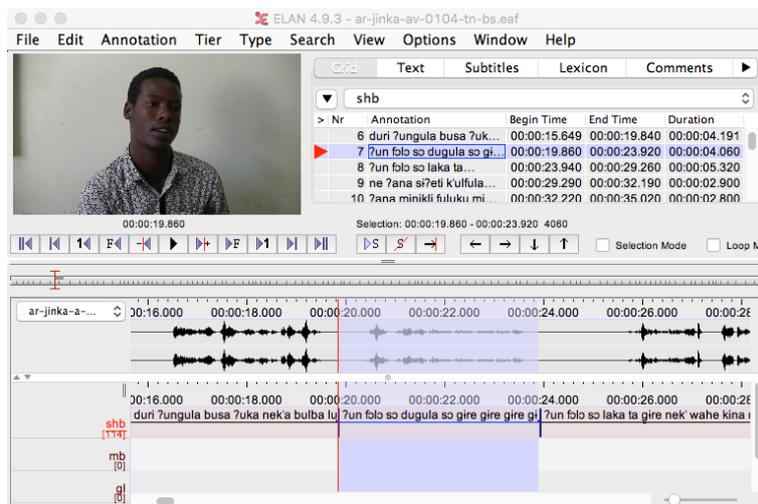


図 1 ELAN (音訳を入力中)

る機能は備わっていない。ELAN が作成するのはあくまで注釈テキストであるが、それを多様な形式に変換・出力することは可能である。たとえば、Toolbox (Field Linguistics Toolbox), FLEx (Fieldworks Language Explorer), Praat など各種ツール用の形式や、SRT (SubRip Text) などの字幕形式として出力することはできる。

SRT 形式は近年の動画提供ウェブサイトでも利用可能であることが多い。そのため、動画提供サイトをプラットフォームとした「字幕付き動画」教材、という方向性も考えられる。だが、動画教材としては単に「字幕付き動画」を再生するだけでなく、たとえば任意の語句が使われている場面を繰り返し再生するなど、字幕 (注釈テキスト) と動画とが双方向に結び付くようなインターフェイスの方がより望ましい。

1.3 IIIF

IIIF (International Image Interoperability Framework) に目を向けると、Presentation API バージョン 3.0 (2020 年 6 月 3 日) から音声と動画をより適切に扱えるようにするタイムラインが導入された [4]。また、IIIF Cookbook には動画に字幕を付与する Manifest の例も掲載されている [5]。つまり IIIF ビューワのインターフェイスを利用して「字幕付き動画」を提供する、という方向性が現実的なものとなった*1。

ただ 2021 年末の時点で、Presentation API 3.0 および動画の読み込みの両方に対応している IIIF ビューワには Universal Viewer と Mirador が存在していたものの [7], [8], どちらにも動画上に字幕を容易に表示する機能は無かった。

Mirador での動画アノテーションに関しては、東京大学大学院人文社会系研究科附属次世代人文学開発センター

人文学部、一般財団法人人文学研究所、東京外国語大学アジア・アフリカ言語文化研究所、有限会社フェリックス・スタイルにより、「Mirador 動画アノテーション対応版」が 2020 年 12 月に公開されている [9]。これは、注釈テキストに対応する動画上の領域を明瞭に視覚化することを実現したものである。

今回、筆者らは「Mirador 動画アノテーション対応版」を改良し、字幕の表示、および言語学習を想定したいくつかの機能追加を行った。

1.4 目的

「Mirador 動画アノテーション対応版」の改良を通し、目指す成果は次のとおりである。

- 学習者が特別なツールを導入することなく、一貫したインターフェイスで字幕付き動画を操作できるプラットフォームを構築する。
- ELAN の注釈データを IIIF で利用するための方法論を構築する。

2. Mirador と ELAN の連携

2.1 Mirador

「Mirador 動画アノテーション対応版」の改良によって追加された機能について、具体的な箇所を図 2 に示す。

1. 注釈サイドバーの自動スクロール

Mirador は注釈テキストをサイドバーに表示する。そこで、動画の再生時間に応じてサイドバーの注釈テキストを自動選択する機能を追加した。表示域の外にある注釈が選択された場合はその注釈がサイドバー中央に来よう自動スクロールされる。

*1 ELAN と IIIF の連携の試みには Graf et al. (2019) がある [6]。これは Image API 2.1 を動画用に拡張するものであり、本稿の目的とは異なるが注目すべき研究である。

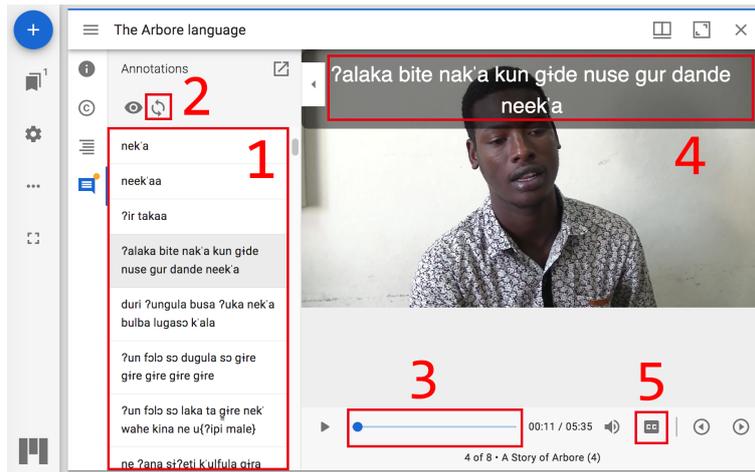


図 2 Mirador に追加された機能

2. 自動スクロールのオン・オフ

サイドバーの注釈テキスト一覧から特定の語句を探すような場合は自動スクロールがかえって妨げとなる．そこで自動スクロールのオン・オフを切り替えるボタンをサイドバーのヘッダ領域に設けた．

3. サイドバーとシークバーの双方向の連携

シークバーから再生箇所を選択したとき、その時間帯に対応する注釈テキストまでサイドバーが自動スクロールするよう機能追加を行った．すなわち、サイドバーの注釈テキストから再生箇所へ飛びだけでなく、動画の再生箇所から注釈テキストへ飛びという双方向の連携が可能になった．

4. WebVTT による字幕

Mirador は内部的に HTML の video 要素を生成する．また近年のウェブブラウザは track 要素を通して WebVTT などの字幕形式を利用することができる [10]．そこで Mirador に WebVTT ファイルを読み込ませ、動画上に字幕を表示する機能を追加した．

WebVTT は仕様上、スタイル記述によって字幕の位置やフォントなどを比較的柔軟に制御することが可能である (図 3)．ただし、実際にどのようなスタイル記述が可能であるかはウェブブラウザのサポート状況に依存する．

5. 字幕のオン・オフ

WebVTT による字幕のオン・オフを切り替えるボタンをシークバーの隣に設けた．なお現時点では音訳、翻訳等の字幕を個別に切り替えることはできない．

2.2 字幕表示のための IIIF 記述

前節に挙げた改良項目から示唆されるように、Mirador のサイドバーに表示される注釈テキストと、WebVTT に

表 1 motivation プロパティの値と意味

値	意味
painting	Canvas を構成するもの．
supplementing	Canvas に関連するもの．
その他	他の規格・仕様で定められたもの．

よって動画上に表示される字幕テキストとは、異なるリソースとして用意する必要がある．本節では具体的な IIIF 記述について説明するが、その前に Presentation API 3.0 のデータモデルを確認しておきたい．

Presentation API 3.0 は以前のバージョンとはデータモデルがやや異なっており、以下のような特徴を持つ [4]．

- Manifest は複数の Canvas を持ちうる．
- Canvas は複数の AnnotationPage を持ちうる．AnnotationPage は常に Canvas 上に置かれるもの (items プロパティに格納される) と、必要に応じて利用されるもの (annotations プロパティに格納されるもの) とに分けられる．
- AnnotationPage は複数の Annotation を持ちうる．
- Annotation は実際のリソースを内包あるいは参照する．また motivation プロパティに表 1 のような値を持つことができる．

以上を踏まえると、動画、字幕テキスト (WebVTT ファイル)、注釈テキストは次のように記述することになる．

Annotation	motivation	Canvas
動画	painting	items
WebVTT	supplementing	annotations
注釈テキスト	commenting	annotations

具体的には、Canvas の構造は次のとおりである．

```

1 { "type" : "Canvas",
2   "id" : "https://example.org/iiif/canvas/1"
3   "items" : [

```

```
1 00:00:06.697 --> 00:00:09.131 line:0
2 neek'aa
3
4 00:00:06.697 --> 00:00:09.131 line:55%
5 <i>The lion</i>
```

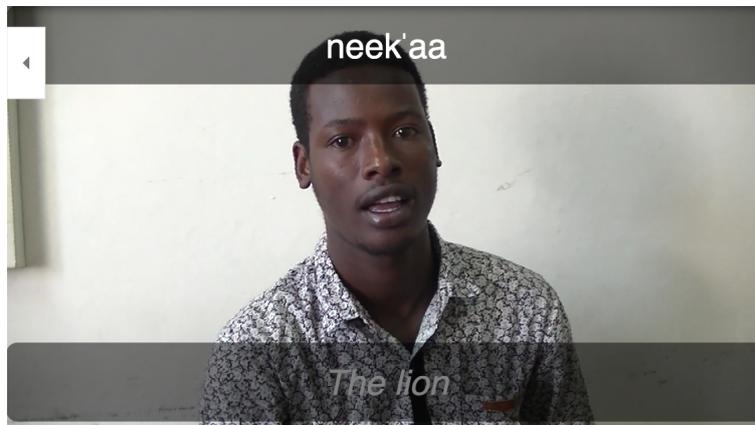


図 3 WebVTT 記述と動画上の字幕（音訳と翻訳）

```
4 { "type" : "AnnotationPage",
5   "items" : [ 動画 Annotation ]
6 }
7 ],
8 "annotations" : [
9   { "type" : "AnnotationPage",
10    "items" : [ WebVTT Annotation ]
11   },
12   { "type" : "AnnotationPage",
13    "items" : [ 注釈テキスト Annotation ]
14   }
15 ]
16 }
```

動画 Annotation の記述は次のとおりである .

```
1 { "type" : "Annotation",
2   "motivation" : "painting",
3   "body" : {
4     "id" : 動画-URI,
5     "type" : "Video",
6     "format" : "video/mp4",
7   },
8   "target" : Canvas-URI#t=0, 再生時間
9 }
```

WebVTT Annotation の記述は次のとおりである . なお、複数の WebVTT ファイルが参照された場合は先頭のみ読み込む .

```
1 { "type" : "Annotation",
2   "motivation" : "supplementing",
3   "body" : {
4     "id" : WebVTT ファイル-URI,
5     "type" : "Text",
6     "format" : "text/vtt",
```

```
7 },
8   "target" : Canvas-URI
9 }
```

注釈テキスト Annotation の記述は以下のとおりである . 1 つの注釈テキストにつき 1 つの Annotation を設けるため、注釈テキストの数だけ同様の Annotation を用意する .

```
1 { "type" : "Annotation",
2   "motivation" : "commenting",
3   "body" : {
4     "type" : "TextualBody",
5     "value" : 注釈テキスト
6   },
7   "target" : Canvas-URI#t= 開始秒, 終了秒
8 }
```

以上、動画上に字幕を表示させるための WebVTT ファイルと、注釈テキストをサイドバーに表示させるための IIIF 記述について説明した . 完全な IIIF Manifest としては The Ethiopian Language Archive に置かれたファイルを参照されたい [11] .

2.3 注釈テキストのエコシステム

本節では ELAN の注釈データを抽出し、IIIF をはじめとする各形式に変換する方法を説明する .

2.3.1 EAF→IIIF, WebVTT

ELAN の注釈データは EAF (ELAN Annotation Format) と呼ばれる XML 形式で保存されている [12] . EAF の仕様は公開されており、主要部分のみ抜粋すると次のような構造をしている *2 .

*2 視認性のために敢えて空白を入れたため、例示のテキストデータ

```

1 <ANNOTATION_DOCUMENT ...>
2   <HEADER TIME_UNITS="milliseconds" ...>
3     ...
4   </HEADER>
5   <TIME_ORDER>
6     <TIME_SLOT TIME_SLOT_ID="ts5"
7       TIME_VALUE="6697" />
8     <TIME_SLOT TIME_SLOT_ID="ts6"
9       TIME_VALUE="9131" />
10    ...
11  </TIME_ORDER>
12  <TIER TIER_ID="shb" ...>
13    <ANNOTATION>
14      <ALIGNABLE_ANNOTATION
15        TIME_SLOT_REF1="ts5"
16        TIME_SLOT_REF2="ts6">
17        <ANNOTATION_VALUE>
18          neek'aa
19        </ANNOTATION_VALUE>
20      </ALIGNABLE_ANNOTATION>
21    </ANNOTATION>
22    ...
23  </TIER>
24  ...
25 </ANNOTATION_DOCUMENT>

```

注目すべきは次の点である。

- 注釈テキストに直接時間帯を埋め込むのではなく、時間ポイントの定義と注釈テキストとが分離されている。
- 時間ポイントは TIME_ORDER 要素内の TIME_SLOT 要素によって定義され、必ず ID を持つ。
- 注釈テキストは ALIGNABLE_ANNOTATION 要素の中に置かれる。また、開始時間を表す @TIME_SLOT_REF1、終了時間を表す @TIME_SLOT_REF2 はそれぞれの時間ポイントを ID 参照する。例では注釈テキスト neek'aa が、6,697 ミリ秒から 9,131 ミリ秒の時間帯に属することを表している。
- 注釈テキストは TIER 要素によってグループ化される。したがって、IIIF の注釈テキスト Annotation へは次のように変換すれば良い。Canvas 上の時間指定 t= は秒で指定することに注意されたい。

```

1 { "type" : "Annotation",
2   "motivation" : "commenting",
3   "body" : {
4     "type" : "TextualBody",
5     "value" : "nee'kaa"
6   },
7   "target" : "Canvas-URI#t=6.697,9.139"
8 }

```

また、WebVTT 形式への変換も容易である。

前後には、実際にはない空白が入っていることに注意されたい。

```

1 WEBVTT
2
3 shb-1
4 00:00:06.697 --> 00:00:09.139
5 nee'kaa

```

2.3.2 EAF→TEI

EAF 形式では時間ポイントの定義と注釈テキストとが分離されている。これにより、同じ時間ポイントを参照する複数の注釈テキストは、もし実際の時間値が変更されたとしても「同期」される。

同じような構造を採用しているのが、TEI (Text Encoding Initiative) 形式における timeline 要素と when 要素である [13]。時間ポイントは when 要素によって定義される。また注釈テキスト (下記の例では u 要素で表現) は開始時間と終了時間を ID で参照する。

したがって、EAF 形式から TEI 形式への変換は比較的容易である。

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0" ...>
2   ...
3   <text>
4     <body>
5       <timeline unit="ms">
6         <when xml:id="ts5"
7           absolute="00:00:06.697"/>
8         <when xml:id="ts6"
9           absolute="00:00:09.139"/>
10        ...
11      </timeline>
12      <annotationBlock>
13        <u start="#ts5" end="#ts6">
14          nee'kaa
15        </u>
16        ...
17      </annotationBlock>
18    </body>
19  </text>
20 </TEI>

```

2.3.3 TTML→WebVTT

TTML (Timed Text Markup Language) は SMIL (Synchronized Multimedia Integration Language) から派生した字幕の規格である *3。

TTML は多様な表現が可能であるが、最も単純なのは次のように時間値を埋め込んだ注釈テキストを並べた形で

*3 2009 年までは TT AF (Timed Text Authoring Format) の DFXP (Distribution Format Exchange Profile) と呼ばれていたが、2010 年から TTML に名称が変更され、2018 年には TTML2 が W3C 勧告となった [14]。一方、HTML Living Standard を推進している WHATWG は 2010 年、SRT 形式を基盤とする簡易な字幕規格を WebSRT として提案し、後に WebVTT として W3C の勧告策定プロセスに載せた。WebVTT は現在は勧告候補である [10]。

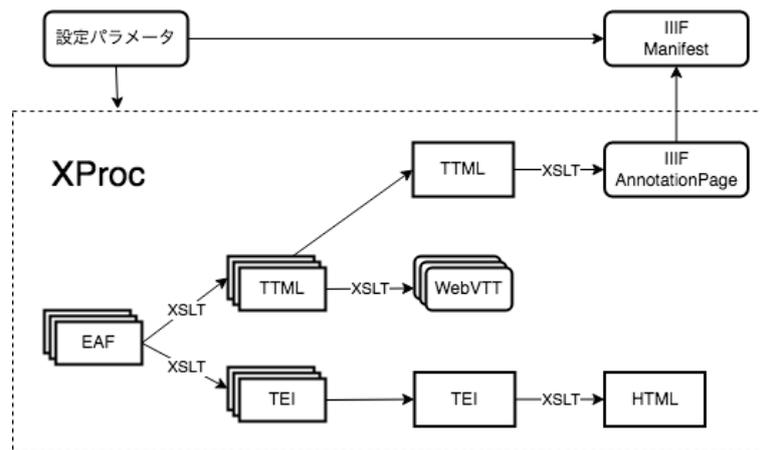


図 4 XML パイプライン

ある．言わば WebVTT 形式をほぼそのまま XML 化したような形になる．

```

1 <tt xmlns="http://www.w3.org/ns/ttml">
2   ...
3   <body>
4     <div region="shb">
5       <p begin="00:00:06.697"
6         end="00:00:09.670">
7         nee'kaa
8       </p>
9       ...
10    </div>
11  </body>
12 </tt>

```

したがって、EAF 形式から TTML 形式に変換し、それから WebVTT 形式に変換すれば、TTML と WebVTT の両方の形式のデータを得ることができる。

2.3.4 XML パイプライン

前節まで見てきたように、EAF 形式は IIIF 形式、WebVTT 形式へと比較的容易に変換可能であるのみならず、その過程において TEI 形式、TTML 形式など多くの副産物を得ることができる。言わば、多様なプラットフォームを見据えた「注釈テキストのエコシステム」の構築が可能だと思われる。

今回は「エコシステム」の可能性を探るため、次のようなパイプラインを考案した。

- EAF 形式を TTML 形式に変換し、そこから IIIF 形式と WebVTT 形式に変換する。
- EAF 形式を TEI 形式に変換し、HTML 形式に変換する。

より具体的には図 4 のようになる。図における通常の四角形は XML 形式を表し、角丸の四角形は非 XML 形式を表す。四角形が重なっているものは複数のファイルを扱えるステップであることを表す。また、EAF 形式に含まれない情報は「設定パラメータ」として JSON 形式で与えた。

パイプラインの実装にあたっては、大部分が XML に関わるため XProc (XML Pipeline Language) を採用し、実行には XML Calabash (バージョン 1.4.1-100) を使用した [15], [16]*4。また、非 XML の領域および XProc/Calabash との連携部分、すなわち実質的な起動スクリプトを JavaScript で記述し、Node.js (バージョン 15.14.0) で実行した [18]。

以上のパイプライン・スクリプトにより、複数の EAF ファイルをまとめて 1 つの IIIF Manifest に変換することに加え、様々な形式の字幕データや、人間可読な音訳、翻訳等のテキストを同時に生成することが可能になった。

3. おわりに

3.1 達成

「Mirador 動画アノテーション対応版」の改良を通して次のような成果が得られた。

- Mirador に自動スクロール機能の追加、サイドバーとシークバーの連携強化、字幕表示機能の追加を行った。
- 学習者は Mirador の一貫したインターフェイスを用い、動画と字幕テキストの制御を直感的に行うことが可能になった。
- ELAN で作成された注釈データ (EAF 形式) から、IIIF Manifest、および様々な形式の字幕データ、人間可読な電子テキストを同時に生成することが可能になった。これは、多様なプラットフォームを見据えた「注釈テキストのエコシステム」が可能であることを裏付けるものである。

3.2 今後の展望

本稿で論じた「Mirador 動画アノテーション対応版」の改良版については、JAEL のウェブページに情報を掲載している [19]。また、パイプライン・スクリプトによって生

*4 XML Calabash は Oxygen XML Editor にも組み込まれている [17]。

成された各種データと動画は、The Ethiopian Language Archive で利用することができる [11]。このウェブサイトは試験運用中であり、今後改めて大量の動画に耐えるウェブサイトの構築を進める予定である。

パイプライン・スクリプトも上記のページから得られる。現状のスクリプトはコマンドライン実行が必要であり、学習コストの高さが課題であるが、今後は上述したウェブサイト構築と連携し、動画と EAF ファイルのアップロード機能、ダウンロード機能を追加する予定である。

謝辞 本研究は JSPS 科研費 18KK0009(2018～2022 年度国際共同研究加速基金(国際共同研究強化(B))「エチオピア諸語の記述とドキュメンテーション: ソーシャル・イノベーションにむけて」代表: 乾秀行, 山口大学)の助成を受けたものである。

参考文献

- [1] Eberhard, D. M., Simons, G. F., and Fennig, C. D.(eds.): *Ethnologue: Languages of the World*, SIL International, Dallas, Texas, twenty-fifth edition (2022). 入手先 <https://www.ethnologue.com/country/ET>(参照 2022-04-28)。
- [2] Japan Association of Ethiopian Languages: Studies in Ethiopian Languages, JAEL (online). 入手先 <https://jael.info/SEL/> (参照 2022-04-28)。
- [3] The Language Archive: ELAN, Max Planck Institute for Psycholinguistics (online). 入手先 <https://archive.mpi.nl/tla/elan> (参照 2022-04-28)。
- [4] Appleby, M., Crane, T., Sanderson, R., Stroop, J. and Warner, S.: IIF Presentation API 3.0, IIF (online) (2020). 入手先 <https://iif.io/api/presentation/3.0/>(参照 2022-04-28)。
- [5] IIF Community: Using Caption and Subtitle Files with Video Content, IIF Cookbook (online). 入手先 <https://iif.io/api/cookbook/recipe/0219-using-caption-file/> (参照 2022-04-28)。
- [6] Graf, J., Rau, F. and Blumtritt, J.: Optimizing Interoperability of Language Resources with the Upcoming IIF AV Specifications, *Proceedings CLARIN Annual Conference 2019* (Simov, K. and Eskevich, M., eds.), Leipzig, Germany, pp. 142–145 (2019). 入手先 <https://www.clarin.eu/event/2019/clarin-annual-conference-2019-leipzig-germany>。
- [7] Universal Viewer: Universal Viewer, 入手先 <http://universalviewer.io/> (参照 2022-04-28)。
- [8] Mirador: Mirador, 入手先 <https://projectmirador.org/> (参照 2022-04-28)。
- [9] 東京大学人文情報学: IIF 動画アノテーション, 東京大学人文情報学(オンライン)。入手先 <https://dh.l.u-tokyo.ac.jp/activity/iif/video-annotation> (参照 2022-04-28)。
- [10] Pfeiffer, S.: WebVTT: The Web Video Text Tracks Format, W3C Candidate Recommendation (online) (2019). 入手先 <https://www.w3.org/TR/2019/CR-webvtt1-20190404/> (参照 2022-04-28)。
- [11] Japan Association of Ethiopian Languages: The Ethiopian Language Archive, JAEL (online). 入手先 <https://dev.jael.info/documentation/> (参照 2022-04-28)。
- [12] The Language Archive: ELAN Annotation Format: Schema version 3.0, Max Planck Institute for Psycholinguistics (online) (2017). 入手先 https://www.mpi.nl/tools/elan/EAF_Annotation_Format_3.0_and_ELAN.pdf (参照 2022-04-28)。
- [13] Text Encoding Initiative: Transcriptions of Speech, P5: Guidelines for Electronic Text Encoding and Interchange, Version 4.4.0 (online) (2022). 入手先 <https://tei-c.org/release/doc/tei-p5-doc/en/html/TS.html> (参照 2022-04-28)。
- [14] Adams, G. and Concolato, C.: Timed Text Markup Language 2 (TTML2), W3C Recommendation (online) (2018). 入手先 <https://www.w3.org/TR/2018/REC-ttml2-20181108/> (参照 2022-04-28)。
- [15] Walsh, N., Milowski, A. and Thompson, H. S.: XProc: An XML Pipeline Language, W3C Recommendation (online) (2010). 入手先 <http://www.w3.org/TR/2010/REC-xproc-20100511/> (参照 2022-04-28)。
- [16] Walsh, N.: Welcome to XML Calabash (2020). 入手先 <https://xmlcalabash.com/> (参照 2022-04-28)。
- [17] SyncRO Soft SRL: XProc Support. Oxygen XML Editor, 入手先 https://www.oxygenxml.com/xml_editor/xproc.html (参照 2022-04-28)。
- [18] OpenJS Foundation: Node.js, OpenJS Foundation (online). 入手先 <https://nodejs.org/> (参照 2022-04-28)。
- [19] Japan Association of Ethiopian Languages: Mirador 3 動画アノテーション対応版, JAEL (online). 入手先 <https://dev.jael.info/mirador3va/> (参照 2022-04-28)。