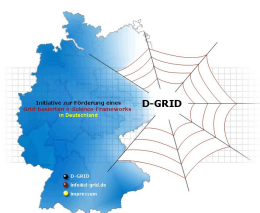


# Linking

Version November 2006  
Arbeitspaket AP 1, Report 1.2  
verantwortlicher Partner: FH Worms

## TextGrid

wissenschaftliche Textdatenverarbeitung -  
ein Community-Grid für die Geisteswissenschaften



Bundesministerium  
für Bildung  
und Forschung

Projekt: **TextGrid**

Teil des D-Grid Verbundes und der deutschen e-Science Initiative

BMBF Förderkennzeichen: 07TG01A-H

Laufzeit: Februar 2006 - Januar 2009

Dokumentstatus: final

Verfügbarkeit: öffentlich

Autoren:

Christoph Ludwig, FH Worms

## Inhaltsübersicht

1. Einleitung.....	4
2. Textannotation.....	4
2.1 Metadateneditor.....	4
2.2 Erklärungsapparate, Werkkommentar etc.....	5
2.3 Normdaten.....	6
2.4 Interne Querverweise.....	7
2.5 Varianten.....	7
3. Manuskriptannotation.....	8
3.1 Physikalische Beschreibung.....	8
3.2 Transkription.....	8
3.3 Faksimile-Annotation.....	9
4. Über TextGrid hinausgehende Link-Editoren.....	10
4.1 Bildwissenschaftliche Annotationen.....	10
4.2 Soundannotation.....	10
4.3 Filmannotation.....	11
5. Bibliographie-Verwaltung.....	11
Anhang A: Bibliographie.....	11

# 1. Einleitung

## 2. Textannotation

Editionsphilologische Arbeiten gehen weit darüber hinaus, (mitunter umfangreiche) Texte zur Verfügung zu stellen: Vielmehr sollen die Texte für die Leser so erschlossen werden, dass die Entstehungs-, Überlieferungs- und Rezeptionsgeschichte nachvollziehbar wird, sachliche und historische Querverbindungen im Inhalt erkennbar werden usw. Ähnliches gilt für andere sprachwissenschaftliche Disziplinen wie der Lexikographie, die beispielsweise Verbindungen zwischen Wörterbüchern herstellt, die den Sprachgebrauch in zeitlich, geographisch oder sozial verschiedenen Umfeldern wiedergeben.

Hierzu müssen die Texte mit Zusatzinformationen, internen und externen Querverweisen sowie bibliographischen Daten angereichert werden. Weil die mit den in TextGrid bereitgestellten Mitteln edierten Texte letztlich als XML-Daten vorliegen werden, die einem Schema gemäß TEI [?] genügen, ist prinzipiell ein generischer XML-Editor hinreichend, um die Texte mit beliebigen Annotationen anzureichern. In der Praxis ist so aber kein effizientes Arbeiten möglich; vielmehr werden für die verschiedenen Arten von Zusatzinformationen spezialisierte Werkzeuge und Dienste mit jeweils geeigneten Nutzerschnittstellen benötigt, um die in nicht-trivialen Projekten anfallenden Daten zu verwalten und adäquat zu präsentieren.

Wir werden deshalb im Folgenden die Werkzeuge und Dienste getrennt betrachten, die Teil der von TextGrid bereitgestellten Workbench sein müssen, um Metadaten, Erklärungen, Normdatenreferenzen, Querverweise innerhalb eines Textes und Referenzen zwischen verschiedenen Varianten eines Textes effizient bearbeiten zu können.

### 2.1 Metadateneditor

Wir verstehen hier unter *Metadaten* strukturierte Informationen, die ein Dokument (bzw. seinen Inhalt) als ganzes beschreiben. In anderen Bereichen wird dieser Begriff auch umfassender gebraucht: Beispielsweise meint *Metadata Extraction* im Rahmen des EARS Projektes [?] die Klassifizierung von einzelnen Textpassagen in einer Transkription. Derlei fällt in TextGrid eher in den Bereich des TEI-Markups.

Typische Metadaten sind z. B. die bibliographischen Angaben zu einem Text wie Autor, Titel, Datum der Veröffentlichung usw. Auch archivarisches Informationen (in welchem Archiv liegt das Originalmanuskript?), Informationen zur Bearbeitungsgeschichte eines Dokuments innerhalb TextGrid (von welchem Dienst und mit welchen Parametern wurde das Dokument erzeugt?) und Schlagworte zur Beschreibung des Inhalts sind einschlägig.

Endgültige Aussagen über die in TextGrid benötigten Möglichkeiten zur Verarbeitung von Metadaten werden erst möglich sein, wenn feststeht, welche Metadaten in TextGrid obligatorisch oder optional erhoben werden. Werkzeuge, die Dublin Core [?] Metadaten erfassen, werden aber mit Sicherheit als Vorlage dienen können. Ein Beispiel hierfür ist der Reggie Metadata Editor [<http://metadata.net/dstc/index.html>]. Auch wenn dieses Tool für die Zwecke von TextGrid zu rudimentär ist, weil es ausschließlich den Export von jeweils neu einzugebenden Daten erlaubt und keinerlei Unterstützung für die Konsistenzüberprüfung etc. bietet, wird dennoch häufig eine ähnlich einfache Nutzerschnittstelle ausreichend sein. Es bietet sich deshalb an, dass TextGrid in einem ersten Schritt einen Metadateneditor implementiert, dessen Interface an Reggie angelehnt ist. Allerdings muss dieser Editor in TextGrids Eclipse Plattform eingebettet werden, Felder vorbelegen, wenn sich sinnvolle Vorschläge aus den bereits in TextGrid vorhandenen Informationen ableiten lassen. Ferner muss der Editor die Angaben auf Ihre Vollständigkeit und syntaktische Korrektheit hin validieren und die Konsistenz z. B. mit den Daten in den TEI- oder TIFF-Header der Dokumente garantieren.

Längerfristig ist es wünschenswert, dass TextGrid auch die (halb-)automatische semantische Annotation von Textdokumenten unterstützt, die dann wiederum von TextGrids Retrieval Tool gewinnbringend genutzt werden kann. Hierbei ist an Methoden gedacht, wie sie etwa in dem von der EU im 6. Rahmenprogramm geförderten Projekt SEKT [<http://www.sekt-project.org/>] genutzt werden, um Office-Dateien oder Webseiten mit Ontologien zu verknüpfen. Die von SEKT entwickelten Tools bauen u. a. auf dem GATE-Framework [<http://gate.ac.uk/>] auf. Nicht ganz so mächtig, dafür aber technisch einfacher ist die an der University of Sydney entwickelte Software Metasaur [<http://www.springerlink.com/index/E9WUWJG3XAQDPRV2.pdf>], die Vorschläge für semantisch naheliegende Schlagworte anbietet. Mit solchen Methoden kann die Vorgabe eines fixen Schlagwortkataloges vermieden werden; dennoch bildet sich ein stabiles Netz von Begriffen heraus, über die große Datenbestände effizient durchsucht und navigiert werden können. Dies wird beispielsweise von Web-Diensten wie del.icio.us [<http://del.icio.us/>] demonstriert.

## **2.2 Erklärungsapparate, Werkkommentar etc.**

Um den oben genannten Anforderungen an sprachwissenschaftliche Editionen gerecht zu werden, wird ein Editor den Originaltext mit einem ausführlichen Kommentar versehen. Dieser Kommentar wird zahlreiche Verweise auf Stellen im Werk, aber auch auf die existierende wissenschaftliche Literatur, auf für die Entstehungs- oder Rezeptionsgeschichte bedeutsame Personen usw. enthalten. Umgekehrt kann ein Editor auch den Werktext um Verweise auf Erklärungen im Kommentar oder in einem separaten Erklärungsapparat, Namensverzeichnis o. ä. ergänzen, die heute nicht mehr geläufig Begriffe erläutern, aus der historischen Distanz nicht mehr allgemein verständliche politische Anspielungen entschlüsseln, auf Bezüge zu anderen Werken hinweisen etc.

In elektronischen Editionen – und ganz besonders in TextGrid – bieten sich hier sehr viel weitergehende Möglichkeiten. Dies beginnt mit der Präsentation der Erklärungen: Kurze Erläuterungen wie Übersetzungen von Begriffen in heute gebräuchliche Ausdrücke können etwa als Tooltips eingeblendet werden, wenn die Maus auf den jeweiligen Begriff zeigt. Längere Erklärungen können in Popups auf einen Mausklick hin geöffnet werden. Wenn zu viele Fenster auf dem Bildschirm vermieden werden sollen, dann können alternativ auch mehrere (Unter-)Fenster verwaltet werden; in einem wird der Originaltext dargestellt, in einem anderen wird der Ausschnitt des Erklärungsapparates gezeigt, der sich auf die aktuelle Cursorposition im Originaltext bezieht. All diese Präsentationsvarianten können so kombiniert werden, wie es dem Editor am zweckdienlichsten erscheint.

Wenn die vom Editor angegebenen Referenzen online verfügbar sind, so können diese in elektronischen Editionen unmittelbar verlinkt werden. Sofern urheberrechtlich zulässig, können diese Referenzen auch im Volltext der Edition beigefügt werden – von sehr großen Editionsprojekten abgesehen ist hierfür auf modernen Datenträgern in der Regel mehr als genug Platz. Für Editionen, die in TextGrid publiziert werden, ist im Grid ohnehin ausreichend Speicherkapazität vorhanden. Auf diese Weise lässt sich eine Abhängigkeit von der Pflege existierender URLs durch Dritte vermeiden.

In elektronischen Editionen können auch neue Medien für die Erläuterungen genutzt werden: Tondokumente, in denen die Aussprache von mundartlichen Ausdrücken wiedergegeben wird; qualitativ hochwertige Farbphotographien, deren Reproduktion im Druck aus ökonomischen Gründen nicht möglich ist; Karten, die dem Nutzer verschiedene Detailstufen und Ansichten bieten; all dies und mehr steht nur digital zur Verfügung.

Schließlich wird es mit TextGrid auch möglich sein, dass Leser die Edition kommentieren oder ergänzen. Die Details, wie der Projektverantwortliche festlegen kann, welcher Leser öffentlich sichtbare Ergänzungen vornehmen kann, müssen im Rahmen der TextGrid-VO-Verwaltung definiert werden. Doch auf diese Weise ist grundsätzlich ein kollaboratives Arbeiten schon während, aber auch über die eigentliche Entstehungsphase einer Edition hinaus möglich.

Aus all dem ergibt sich, dass ein Editor für Erläuterungen zunächst einmal Freitext unterstützen muss. Darüber hinaus sollte die Möglichkeit bieten, beliebige Ressourcen über URIs zu verlinken. Um weitergehende Erläuterungen zu Hintergrundinformationen oder Kommentare durch Kollegen zu ermöglichen, sollte der Editor erlauben, Erläuterungen zu verschachteln und untereinander zu vernetzen.

Wie die Erläuterungen einem Leser präsentiert werden, liegt außerhalb des Aufgabenfeldes des Editors für Erläuterungen – dies muss durch die Publishing-Module für das jeweilige Ausgabemedium geeignet definiert werden. Der Erläuterungs-Editor legt lediglich den Inhalt der Erläuterungen und ihre Verknüpfungen fest.

John Bradleys Eclipse Plugin *Pliny* [<http://pliny.cch.kcl.ac.uk/index.html>] implementiert eine elaborierte Notizenverwaltung. Diese Notizen können untereinander sowie mit Webseiten, PDF-Dokumenten und Bildern verknüpft werden. Zwar ist es derzeit nicht möglich, die Notizen direkt in TEI-Dokumente einzubinden; aber auf den ersten Blick scheint ein solches Feature mit einem Aufwand nachzurüsten zu sein, der wesentlich kleiner ist als der Aufwand für eine Neuimplementierung eines vergleichbaren Editor für Erläuterungen von Grund auf. Wir empfehlen deshalb, dass die Architektur und das API von Pliny technisch genauer evaluiert werden mit dem Ziel, Pliny zu einem TextGrid-Editor für Erläuterungen auszubauen.

## 2.3 Normdaten

Eine spezielle Variante von Erläuterungen sind Normdatenreferenzen. Insbesondere in historischen Texten finden sich für Eigennamen, geographische Bezeichnungen etc. zahlreiche verschiedene Schreibweisen. Mitunter gibt es für eine Person oder einen Ort auch unterschiedliche Namen, die nicht einmal in der Aussprache Ähnlichkeiten aufweisen. Ein Beispiel hierfür aus der jüngeren Geschichte ist die Stadt Chemnitz, die in der zweiten Hälfte des 20. Jahrhunderts für knapp 40 Jahre Karl-Marx-Stadt hieß. Selbst eine unscharfe Suche nach „Chemnitz“ in einem Textkorpus wird alle Vorkommen von „Karl-Marx-Stadt“ übergehen, wenn nicht die Städtenamen im Korpus mit einer Normdatenbank verknüpft sind, die einerseits einheitliche Referenznamen festlegt und andererseits mehrdeutige Namen disambiguiert.

Um derlei Referenzen effizient erstellen und verwalten zu können, ist ein spezialisierter Normdaten-Linkeditor notwendig. Dieser erstellt für die einschlägigen Wörter Vorschlagslisten, auf welche Einträge in der Normdatenbank sie sich beziehen, aus denen der Anwender wählen kann. Um dies zu erleichtern, muss der Linkeditor immer auch den Kontext wiedergeben, in dem der Begriff im vorliegenden Text gebraucht wird. Falls in den ausgewählten Normdatenbanken kein passender Eintrag vorhanden ist, dann muss der Anwender in der Lage sein, entsprechende lokalen Referenzlisten geeignet zu ergänzen.

Die Begriffe, die mit einer Normdatenbank verknüpft werden müssen, muss der Editor auf verschiedene Weisen erkennen können. Zunächst können Eigennamen etc. bereits zuvor von einem Lemmatisierer als solche erkannt und geeignet markiert worden sein. Zweitens gibt es die Möglichkeit, den Textkorpus direkt mit den in der Normdatenbank hinterlegten bekannten Varianten abzugleichen und auf diese Weise eine Liste der potentiellen Belegstellen im Korpus zu ermitteln. Schließlich muss es dem Anwender auch immer möglich sein, eine Referenz manuell zu ergänzen.

Es existieren nach unserer Kenntnis derzeit keine Tools, die all diese Möglichkeiten bzw. Kombinationen davon direkt unterstützen. Die vorhandenen Implementierungen von Algorithmen zur Namenserkennung (etwa in GATE [<http://gate.ak.ac>], ESpotter [<http://kmi.open.ac.uk/people/jianhan/ESpotter/>]) variieren z. B. darin, wie sehr sie den weiteren Kontext einziehen. Auch ist die Erkennungsleistung dieser Tools bei historischen Texten zu hinterfragen. Es wird deshalb notwendig sein, den Editor für Normdatenreferenzen in TextGrid zu großen Teilen neu zu entwickeln, wobei aber auf die Algorithmen in den existierenden Werkzeugen zurückgegriffen werden kann.

## 2.4 Interne Querverweise

Die bislang beschriebenen Verknüpfungen bezogen sich durchweg auf Informationen, die außerhalb des edierten Textes stehen. Häufig ist es aber auch wünschenswert, Verbindungen innerhalb eines Textes navigierbar zu machen. In manchen Fällen werden diese Querverweise vom Editor hinzugefügt, um den Leser das Auffinden z. B. von Parallelstellen zu erleichtern. In anderen Fällen, etwa bei der Edition von Wörterbüchern, wird der Autor bereits explizit auf eine andere Textstelle verwiesen haben; hier ist es das Ziel, diese Referenz elektronisch abzubilden. Weil die Texte in TextGrid als TEI-kodierte XML-Dokumente vorliegen werden, werden solche Querverweise in der Regel als XPointer realisiert werden.

Je nach Anwendungsfall werden die internen Querverweise sinnvollerweise über den Editor für Erläuterungen angelegt, etwa wenn der Herausgeber auf eine Parallelstelle aufmerksam machen möchte. Wenn die Verweise schon vom Autor angegeben wurden, wird man die entsprechenden Verknüpfungen schon in den frühen Stadien der Textauszeichnung im XML-Editor erstellen können.

Der XML-Editor kann auch in Projekten mit zahlreichen Verknüpfungen noch erfolgreich eingesetzt werden; beispielsweise werden im *ellexiko*-Projekt am Institut für Deutsche Sprache die Verknüpfungen der Einträge mit dem Editor XMetal gepflegt.

Wenn man mit großen, sehr engmaschig verknüpften Korpora arbeitet, wie es z. B. im Trierer Wörterbuchnetz [<http://woerterbuchnetz.de/>] der Fall ist, dann kann es auch notwendig sein, dass die Verweise in spezielle Datenbanken ausgelagert werden müssen, um eine performante Suche zu ermöglichen. Das Verwalten solcher Linkdatenbanken und der zugehörigen Indizes geht über die Möglichkeiten eines allgemein einsetzbaren Linkeditors hinaus.

Es ist sehr gut vorstellbar, wenn nicht sogar wahrscheinlich, dass beim Einsatz von TextGrid für konkrete Projekte der Bedarf für Werkzeuge zur Verwaltung interner Verweise aufkommt. Zum jetzigen Zeitpunkt scheinen die Anforderungen aber so allgemein, dass sie von anderen TextGrid-Werkzeugen mit abgedeckt werden oder bereits so speziell, dass eine Implementierung erst realistisch ist, sobald ein Grundstock an allgemein nutzbaren Werkzeugen in TextGrid zur Verfügung steht.

## 2.5 Varianten

In großen Korpora existieren noch weitere interne Referenzen, die sich aus der Kollationierung von Textvarianten ergeben. Häufig liegen nicht alle Textpassagen in identischer Reihenfolge vor, größere Textblöcke wurden ergänzt oder gestrichen. Existierende Kollationierungswerkzeuge sind nur eingeschränkt und mit sehr hohem Aufwand in der Lage, solche Passagen zu erkennen.

Deshalb muss ein Herausgeber regelmäßig vor der Kollationierung sog. Aufsetzpunkte festlegen. Dabei handelt es sich um Anfangs- und Endemarkierungen von Textabschnitten, die sich in den einzelnen Textvarianten nicht in größeren Blöcken unterscheiden, sondern nur in der Schreibung der einzelnen Worte oder in Auslassungen einzelner Wendungen etc. Die Kollationierungstools können dann mit vertretbarem Aufwand die Unterschiede zwischen korrespondierenden Textabschnitten ermitteln.

Die Aufsetzpunkte definieren also eine Korrespondenz zwischen Textabschnitten in verschiedenen Varianten und stellen somit ebenfalls eine Verknüpfung dar. Diese Verknüpfungen sind auch über die eigentliche Kollationierung und die spätere Darstellung der Unterschiede auf Wort oder Satzebene hinaus von Bedeutung: Büchners *Woyzeck* ist ein bekanntes Beispiel, in dem sich aus verschiedenen Gründen die vom Autor zuletzt geplante Reihenfolge der Szenen nicht abschließend klären lässt. Aus den Aufsetzpunkten lässt sich ein schematischer Überblick ableiten, welcher Zusammenhang in den überlieferten Entwürfen jeweils gegeben ist.

Das Werkzeug, mit dem in TextGrid die Aufsetzpunkte definiert werden, wird als Teil der Anwenderschnittstelle des Kollationierungsdienstes implementiert werden.



### 3.Manuskriptannotation

Einer der großen Vorteile elektronischer Editionen ist die Möglichkeit, dem Leser Faksimiles der für die Edition genutzten Manuskripte zugänglich zu machen. Außer für einen kleinen Kreis von Experten ist es im Allgemeinen aber sehr schwer, diese Manuskripte zu lesen – möglicherweise weil die Handschrift des Autors nur für Geübte zu entziffern ist oder weil das Manuskript aufgrund von Alterungsprozessen Schaden genommen hat. Der eigentliche Mehrwert für eine Edition entsteht also dadurch, dass die Manuskripte durch eine geeignete Präsentation und Einbettung in die übrige Edition für den Leser zugänglich werden. Um dem Leser die jeweils passenden Bild- bzw. Textausschnitte zu zeigen und ihn andererseits nicht mit mehr Informationen zu überhäufen, als er verarbeiten kann, sind sehr enge Verknüpfungen notwendig, bei deren Auswertung auch immer der vom Leser gerade betrachtete Kontext berücksichtigt werden muss.

In den folgenden Abschnitten betrachten wir drei Arten von Verknüpfungen genauer, die bei der Annotation von Manuskripten berücksichtigt werden müssen: Informationen zur genaueren Beschreibung der physikalischen Beschaffenheit des Originals, Beschreibungen von graphischen Details, die im Faksimile wiedergegeben sind, und schließlich Verweise, die eine exakte Positionierung des transkribierten Texts im Faksimile zulassen.

#### 3.1 Physikalische Beschreibung

Bei der physikalischen Beschreibung einer Manuskriptseite handelt es sich streng genommen um eine Metadatenerfassung wie in Abschnitt 2.1 beschrieben. Dort lagen v. a. Daten im Fokus, die typischerweise größere Einheiten beschreiben – etwa bibliographische Informationen über ganze Bücher. Hier betrachten wir Metadaten, die sich auf die sehr viel kleineren Einheiten einer einzelnen Manuskriptseite beziehen, etwa deren Abmessungen, ihre stoffliche Beschaffenheit, darin enthaltene Wasserzeichen u. ä. Diese Daten werden üblicherweise bei der Digitalisierung erfasst. (Vgl. hierzu die TextGrid Szenarienbeschreibung [<http://tinyurl.com/yk3caqj>].)

Die Beschreibungen von Manuskriptseiten sind innerhalb eines Editionsprojektes typischerweise strukturiert; denn der Projektleiter legt im Vorfeld der Digitalisierung Erfassungsregeln fest, nach denen alle Daten erhoben werden. Auch gibt es Eigenschaften (z. B. Abmessungen), deren Beschreibung nahezu immer in den Erfassungsregeln gefordert werden wird. Doch ist bei den Erfassungsregeln insgesamt mit einer größeren Varianz zu rechnen als etwa bei den bibliographischen Daten. Dies hat Auswirkungen auf den benutzten Metadateneditor, der für diesen Einsatzzweck erlauben muss, dass die Erfassungsregeln des Projektleiters mit geringem Aufwand in Eingabemasken für die Datenerhebung umgesetzt werden können.

In dem Moment, in dem die Seitenbeschreibungen eingegeben werden, existiert nicht notwendig bereits die zugehörige Scan-Datei im System. Statt dessen muss eine eindeutige Seiten-ID mit erhoben werden, über die der Datensatz später – möglichst ohne manuellen Eingriff – dem Bild zugeordnet werden kann. Diese Zuordnung kann über spezielle Informationen im Header der Bilddatei erfolgen, wie sie viele Bildformate (TIFF, JPEG) unterstützen, oder sie wird über den Dateinamen realisiert, unter dem das Bild in TextGrid importiert wird. Jedenfalls muss ein geeigneter Dienst diese Verknüpfung zwischen den Metadaten und dem Bild explizit machen und verifizieren, dass diese Verknüpfungen am Ende des Digitalisierungsprozesses eine Bijektion zwischen den Metadaten und den Bildern realisieren.

#### 3.2 Transkription

Der Leser einer Edition soll die Erkenntnisse und Entscheidungen des Herausgebers nachvollziehen können. Hierzu wird er sich bei elektronischen Editionen, die Scans der Manuskripte umfassen, auf die Faksimiles beziehen. Es ist allerdings speziell bei handschriftlichen Originalen meist alles andere als einfach, die Manuskripte zu lesen; das Dokument kann vergilbt sein, der Text ist vielleicht redigiert und dabei überschrieben worden oder die Handschrift ist schlicht schwer zu entziffern. Für die meisten Nutzer einer Edition wäre es deshalb nicht mit vertretbarem Aufwand möglich, ohne weitere Hilfestellung den Bezug zwischen dem vom Herausgeber rekonstruierten Text und den Faksimiles



herzustellen. Bei gedruckten Originalen mag dieses Problem weniger eklatant sein, aber auch hier ist es für den Leser nicht immer einfach, in Scans von eng bedruckten Seiten den Überblick zu bewahren, speziell wenn eine kleine Frakturschrift benutzt wurde.

Um den Nutzer der Edition zu unterstützen, bietet es sich deshalb an, die Transkription des Textes durchgängig mit den Faksimile zu verknüpfen. Diese Referenzen können dann herangezogen werden, um beispielsweise die Transkription als Tooltip anzuzeigen, sobald der Leser den Mauszeiger über ein Wort im Faksimile bewegt. Wenn die Transkription und das Faksimile in zwei (Unter-)Fenstern nebeneinander gezeigt werden, so kann auch durch eine besondere Hervorhebung die Stelle im Faksimile angezeigt werden, die mit dem Ausschnitt des Transkripts korrespondiert, in dem sich gerade der Textcursor befindet. Ob diese Verknüpfung auf Satz- Wort- oder gar Zeichenebene erfolgt, macht aus technischer Sicht – von der Menge der zu verarbeitenden Daten abgesehen – keinen großen Unterschied; letztlich bleibt es dem Leiter des Editionsprojektes überlassen, welchen Aufwand er im konkreten Fall für angemessen erachtet.

In jedem Fall wird ein Editor für die Verlinkung von Transkriptionen den Bearbeiter möglichst weitgehend unterstützen müssen, um ein ergonomisches Arbeiten zu ermöglichen. Beispielsweise kann der Editor versuchen, zusammenhängend Schriftzüge und somit Wortgrenzen in handschriftlichen Dokumenten zu erkennen; wenn der Bearbeiter die Verknüpfungen Wort für Wort erstellt, dann kann der Editor folglich automatisch einen Vorschlag für den vom nächsten Wort eingenommenen Bildausschnitt unterbreiten. Die dafür notwendigen Algorithmen finden beispielsweise im Bereich der digitalen Bildbearbeitung bereits Verwendung, die automatisierte Maskierung von zusammenhängenden Bildbereichen ist in Programmen wie Adobe Photoshop oder The Gimp seit langem Stand der Technik. Hiermit ist keine hinreichende Erkennungsleistung zu erwarten, um diesen Prozess vollständig zu automatisieren; aber es würde den Bearbeiter bereits wesentlich entlasten, wenn er nicht jedes Wort im Faksimile manuell markieren muss.

Bei gedruckten Vorlagen ist eine vollautomatisierte Hinterlegung des Faksimiles mit Transkriptinformationen vorstellbar. Hier können OCR-Algorithmen zum Einsatz kommen – unabhängig davon, ob der Text mittels OCR oder durch manuelle Transkription erfasst wurde. Diese Algorithmen bestimmen zwangsläufig die von einzelnen Buchstaben und Worten bedeckten Bildausschnitte. Wenn diese Information nicht verworfen wird und das OCR-Ergebnis mit dem bereits fehlerbereinigten Transkript abgeglichen wird, so ist eine verlässlichen Zuordnung von Textelementen zu Bildausschnitten zu erwarten.

### **3.3 Faksimile-Annotation**

Manuskripte enthalten häufig außer dem darauf geschriebenen Text noch weitere Informationen und Elemente. Typische Beispiele sind Abbildungen und Skizzen, aber auch radierte Stellen, Verschmutzungen oder Beschädigungen. All dies kann für die Edition relevante Information beisteuern: Wenn mehrere Manuskriptseiten ähnliche Wasserflecken an der gleichen Stelle aufweisen, dann ist dies beispielsweise ein Indiz dafür, dass diese Seiten irgendwann einmal zusammen gelagert worden sind. Wird ein solches Indiz als Argument für eine These genutzt, so wird man auf die „Belege“ – sprich die Schadstellen in den Manuskriptseiten – verweisen wollen. Ein Werkkommentar wird typischerweise auf die in den Manuskripten enthaltenen Skizzen eingehen und muss folglich mit dem entsprechenden Bildausschnitt verlinkt werden. Umgekehrt wird man von verschmutzten oder fehlenden Teilen der Manuskriptseite häufig auf eine Erläuterung verweisen wollen, weil im Faksimile nicht mehr notwendig erkennbar ist, weshalb der Text an dieser Stelle nicht mehr lesbar ist.

All diese Erläuterungen sind von Ihrer Art nach ähnlich den Verweisen, wie sie in Werkkommentaren und den diversen Erklärungsapparaten eingesetzt werden (vgl. Abschnitt 2.2). Der Unterschied liegt eher im technischen Bereich: Statt auf ein Element in einer TEI-kodierten XML-Datei, das per XPointer referenziert werden kann, oder eine Ressource im Netz, die über einen URL identifiziert wird, muss hier ein Ausschnitt eines Bildes beschrieben werden. In den meisten Fällen wird dieser Bereich die Gestalt eines Rechtecks haben. Andere Flächen sind sowohl bei der Definition durch den Anwender als auch in

der Verwaltung durch das System sehr viel komplexer. Wir empfehlen deshalb, dass der in TextGrid entwickelte Link-Editor für die Faksimile-Annotation zunächst nur rechteckige Bildbereiche unterstützt.

Dass bereits in Abschnitt 2.2 erwähnte Tool Pliny unterstützt auch Verweise auf Bild- und PDF-Dateien. Falls Pliny sich als Ausgangsbasis für die Entwicklung des Editors für die Verknüpfung von Erläuterungen und Kommentaren bewährt, dass ist es naheliegend, auch den entsprechenden graphischen Link-Editor darauf aufzubauen.

## **4.Über TextGrid hinausgehende Link-Editoren**

Es ist kein Grund erkennbar, weshalb TextGrid notwendig auf textwissenschaftliche Anwendungen im engeren Sinne beschränkt bleiben muss. Beispielsweise könnten bei einer Edition des Werkes von Martin Luther die in TextGrid entwickelten Werkzeuge und Dienste ebenso nutzbringend eingesetzt werden wie bei einer Edition der Essais von Michel de Montaigne. Im Falle Luthers ergäben sich aufgrund der zahlreichen von ihm verfassten bzw. umgedichteten Kirchenlieder bereits wieder Überschneidungen mit der Musikwissenschaft. Auch im Bereich der modernen Medien sind viele Anwendungen vorstellbar. Im derzeitigen TextGrid-Projekt sind jedoch keine Ressourcen vorhanden, um auch die für „benachbarte“ Fachrichtungen benötigten Dienste und Werkzeuge bereitzustellen. Wir skizzieren dennoch einzelne Anwendungsfälle für spezialisierte Link-Editoren aus solchen Bereichen, um potentielle längerfristige Entwicklungsrichtungen von TextGrid aufzuzeigen.

### **4.1Bildwissenschaftliche Annotationen**

Viele mittelalterliche Handschriften sind mit sehr aufwändig geschaffenen Bildern illustriert. Diese müssen auch in einer textwissenschaftlichen Edition erfasst und beschrieben werden. Elektronische Editionen erlauben zudem kostengünstig, die derart ausgestalteten Manuskripte als Faksimile wiederzugeben.

Wenn diese Bilder nicht nur als Illustration des Textes, sondern als eigenständige Objekte des kunsthistorischen Interesses betrachtet werden, dann ergeben sich zusätzliche Fragestellungen: Welche Aussagen zur Mal- und Zeichentechnik sind möglich? Was für Farben kamen zum Einsatz? Hat der Maler auf Vorlagen zurückgegriffen? Manche dieser Fragen können beispielsweise beantwortet werden, indem etwa mit Methoden aus der Radiologie übermalte Skizzen sichtbar gemacht werden. Für eine Edition wäre es dann wünschenswert, wenn diese Skizzen dem Manuskript wieder überlagert werden könnten, zusätzlich mutmaßliche Vorlagen eingeblendet würden etc. Hierfür würde ein Link-Editor benötigt, der Überblendungen von Bildern sowie grundlegende Bildbearbeitungsfunktionalitäten wie Drehungen, Spiegelungen, Stauchungen etc. unterstützt.

Entwicklungen in diesem Bereich könnten auch wieder in die Textwissenschaften rückgekoppelt werden, wenn Sie beispielsweise zur Darstellung getilgter Handschriften auf Palimpsesten genutzt werden.

### **4.2Soundannotation**

Seit dem 20. Jahrhundert stehen auch von Schriftstellern häufig Tondokumente zur Verfügung, etwa Radiointerviews oder Aufzeichnungen von Lesungen und Vorträgen. Ein bekanntes Beispiel sind Thomas Manns Radioansprachen während des 2. Weltkrieges aus dem amerikanischen Exil.

Bereits mit den derzeit in TextGrid entwickelten Werkzeugen wird es möglich sein, z. B. aus Erläuterungen heraus auf Sounddateien zu verweisen. Um das Potential solcher Tondokumente für eine Edition ausschöpfen zu können, wäre ein in TextGrid integriertes Werkzeug erforderlich, mit dem man in Tonspuren navigieren, Schnittmarken definieren kann, über die Anfang und Ende eines referenzierten Ausschnitts festgelegt werden, etc.

In einem weiteren Schritt wäre es dann auch wünschenswert, nicht nur auf Tondokumente (bzw. Ausschnitte daraus) verweisen zu können, sondern auch die Tondateien selbst wieder mit Erläuterungen, Verweise auf andere Dokumente usw. anreichern zu können. Dies wäre spätestens dann notwendig, wenn man etwa Editionen der Werke von Künstlern wie Bob Dylan in TextGrid realisieren wollte, in denen die

(Lied-)Texte einer sehr hohen Variabilität unterliegen – sowohl in Ihrer schriftlichen Form als auch in Ihrer Wiedergabe auf der Bühne bei Live-Auftritten – und weder der schriftlichen noch der in Konzertmitschnitten festgehaltenen gesungenen Form ein für das Werk höheres Gewicht beigemessen werden kann.

### 4.3 Filmannotation

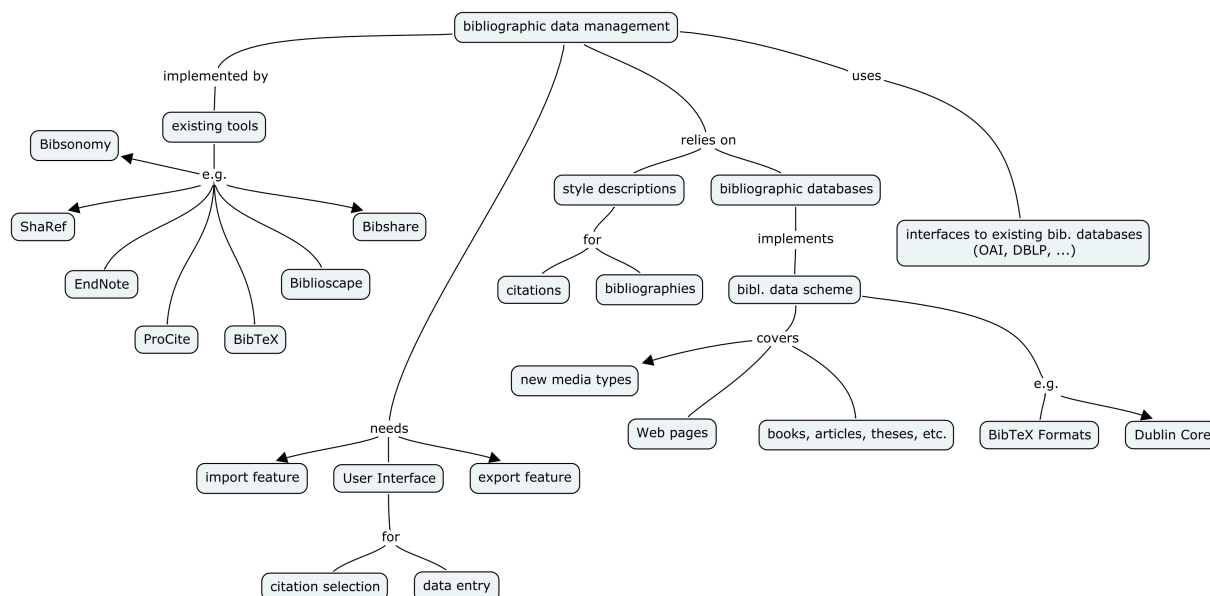
Die Relevanz neuer Medien für die Editionswissenschaften erschöpft sich nicht in dem im letzten Abschnitt skizzierten Bedarf zur Verlinkung mit Tondokumenten. Das gleiche Argument kann auch für Filmdokumente vorgebracht werden. Aufzeichnungen von Schauspielaufführungen, die von den Autoren selbst inszeniert wurden oder die für die Rezeptionsgeschichte bedeutsam waren, können zusätzliche Aspekte zu einer umfassenden Edition beitragen. Bei manchen zeitgeschichtlichen Autoren können sie auch integraler Bestandteil einer vollständigen Werkausgabe sein – beispielsweise hat Samuel Beckett mehrere Filme und Fernsehstücke gedreht.

Auch hier wäre in einem ersten Schritt ein Werkzeug notwendig, das es erlaubt, auf Ausschnitte von Filmen zu verweisen. Für umfangreichere filmwissenschaftliche Arbeiten wären allerdings weitergehende Werkzeuge notwendig, mit denen einzelne Einstellungen oder Szenen annotiert werden könnten. Hierfür wären ferner neue Methoden zur Präsentation und Navigation von Verweise auf derart dynamische Inhalte notwendig.

Derlei Editionsprojekte würden voraussichtlich auch in anderen Bereichen von TextGrid weiteren Entwicklungsbedarf nach sich ziehen: Zwar ist TextGrid prädestiniert, um die bei Bild- und Tondokumenten anfallenden großen Datenmengen zu verwalten; aber bislang ist kein Streaming-Service o. ä. geplant, über den die Clients effizient auf diese Daten zugreifen könnten.

## 5. Bibliographie-Verwaltung

[Existiert bislang leider nur als Concept Map]



## Anhang A: Bibliographie