Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 5345–5355
Marseille, 20-25 June 2022

»textklang« – Towards a Multi-Modal Exploration Platform for German Poetry

Nadja Schauffler*, Toni Bernhart†, André Blessing*, Gunilla Eschenbach‡, Markus Gärtner*, Kerstin Jung*, Anna Kinder‡, Julia Koch*, Sandra Richter‡,†, Gabriel Viehhauser†, Thang Vu*, Lorenz Wesemann‡, Jonas Kuhn*

*Institute for Natural Language Processing,
University of Stuttgart, Germany
{firstname.lastname}@ims.uni-stuttgart.de

†Institute of Literary Studies
University of Stuttgart, Germany
{firstname.lastname}@ilw.uni-stuttgart.de

‡German Literature Archive
Marbach, Germany
{firstname.lastname}@dla-marbach.de

Abstract

We present the steps taken towards an exploration platform for a multi-modal corpus of German lyric poetry from the Romantic era developed in the project »textklang«. This interdisciplinary project develops a mixed-methods approach for systematic investigations of the relationship between written text (here lyric poetry) and its potential and actual sonic realisation (in recitations and musical performances). The multi-modal »textklang« platform will be designed to technically and analytically combine three modalities: the poetic text, the audio signal of a recorded recitation and, at a later stage, music scores of a musical setting of a poem. The methodological workflow will enable scholars to develop hypotheses about the relationship between textual form and sonic/prosodic realisation based on theoretical considerations, text interpretation and evidence from recorded recitations. The full workflow will support hypothesis testing either through systematic corpus analysis alone or with additional contrastive perception experiments. For the experimental track, researchers will be enabled to manipulate prosodic parameters in (re-)synthesised variants of the original recordings. The focus of this paper is on the design of the base corpus and on tools for systematic exploration – placing special emphasis on our response to challenges stemming from multi-modality and the methodologically diverse interdisciplinary setup.

Keywords: Digital Humanities, Literary studies, Lyric poetry, Corpus exploration platform, Speech resource, Prosody

1. Introduction

In this paper we present the creation of a multi-modal platform for research on literature recitations which is being developed within the »textklang« project. »textklang« is an interdisciplinary endeavour targeting the interplay of text and sound in lyric poetry. It has set out to combine methods and disciplinary practices from literary studies, digital humanities/computational linguistics and linguistics/laboratory phonology. The core of the project's corpus is based on German poetry from the Romantic era (approximately 1795–1835) and comprises a collection of recordings of lyric poetry recitations and vocal performances of songs setting Romantic poetry to music. In the course of the project, we will develop a methodological toolbox to investigate the text and sound dimension in this data by combining prosodic analysis with text-analytical tools from corpus linguistics and computational literary studies. Methodology and tools, such as an exploration platform, will enable scholars to generate and test hypotheses about systematic patterns in the relationship between textual form, prosodic realisation and aspects of interpretation. In order to evaluate and test hypotheses regarding the interplay of prosodic realisation and interpretation, prosodic details will be manipulated by means of controlled (re-)synthesised variants of the original recordings and tested in perception experiments. The present paper focuses on the base corpus and tools for systematic multi-modal exploration.

1.1. Motivation

Research in the digital humanities has recently put considerable emphasis on method development for 'scalable' text analysis, supporting processes of 'macroanalysis' of longer texts or entire text corpora (Jockers, 2013). So far, most computational methods have exclusively taken the textual form of the objects of study as the basis for data-driven analysis (which in the digital humanities has to be interlocked with processes of theory-driven reflection (Reiter and Pichler, 2020; Kuhn, 2019)).[1] Few would object that a more com-

[1]Of course, computational work focusing on the analysis of specific aspects of poetry, such as metrical analysis or rhyme pattern detection (for instance Hayes et al. (2012),

prehensive scholarly approach to text should take other dimensions into consideration. In particular, the interplay between textual form, the content level, and the sonic level of poetic text can play a crucial role for an understanding of developments in literary history or in the grounding of theories of literary interpretation.

Developments in speech technology and linguistic research into the interplay of prosody and other linguistic levels has led to an inventory of models and descriptive devices that seem to lend themselves as building blocks for a more comprehensive approach to theoretically grounded data-oriented digital humanities research on texts in their written shape as well as their sonic realisation. A methodological goal of the »textklang« project is to explore to what extent a combination of computational models of text and prosody can productively inform literary research. Poetry from the Romantic era, during which the sonic dimension of language was viewed as particularly central, provides an ideal pilot scenario for a more general methodological endeavour.

We build on prior work on corpus annotation and analysis on both text and speech data (see Section 4). As a major tool for corpus exploration, we use and adapt ICARUS (Gärtner, M. and Thiele, G., 2021), a multimodal exploration and visualisation tool.

1.2. Outline

In this paper we present the steps taken towards this multi-modal exploration platform, from selecting the speech material and digitising it (Section 3.1), recovering the respective text material (Section 3.2), collecting and modelling metadata (Sections 3.3 and 5), automatic annotations of both the text and the speech material (Section 4), to visualising and exploring the created corpus based on an exemplary research question (Section 6). A special focus in this paper is on challenges stemming from the multi-modal and cross-disciplinary setup and our approaches to solving these issues. We hope that at a methodological level, our discussion will be informative for other projects working with a multi-level annotated corpus of text and speech – potentially in a quite different research context.

2. The »textklang« project

The »textklang« project is a collaboration between the German Literature Archive (*Deutsches Literaturarchiv*, DLA) Marbach and several research groups from the University of Stuttgart, combining expertise in quite distinct relevant fields of study – most centrally literary studies, digital humanities, (text-oriented) computational linguistics, laboratory phonology and speech technology (specifically speech synthesis). The project aims to address the relationship between poetic text and its sonic realisation by technically and analytically combining three modalities: text (captured via a digital textual encoding of typeset poems), music scores of musical settings of a poem (captured in MIDI), and the audio signal of a recorded performance – either a recitation of a poem or a vocal performance of a musical setting. The technical treatment of musical settings will be addressed at a later project stage.

The technical platform envisioned in »textklang« is designed to support both exploratory and hypothesis-driven research (by literary scholars or digital humanists) on the multi-modal corpus. We may for example imagine a scholar interested in the phenomenon of enjambment (the continuation of a sentence or phrase from one line of poetry to the next, see also Section 6.1): enjambment creates a tension between the flow of the textual content and the structural form of the poem. In reciting a poem with an enjambment, there are different possibilities rendering the line break in prosodic terms. Comparing actual recitations from a corpus may shed light on systematic patterns – possibly relative to constraining contextual factors (e.g. restricting attention to certain poets or to recitations from a particular period). To track down such patterns, the researcher (who may or may not start out with prior expectations) should be able to explore the corpus using metadata along with various analytical tools which address both the textual and the phonetic level. The findings in a first selection of concrete corpus instances may lead the researcher to formulate a general hypothesis that can be in principle tested against (text and audio) data. (For the sake of illustration, let us say the hypothesis is that enjambments going along with a semantic contrast in the split-up parts of the sentence tend to be marked with tonal means, whereas a semantically continuous sentence running across an enjambment tend to be "smoothed out" prosodically.)

Given the multitude of interacting linguistic and contextual factors, it is however likely that a quantitative analysis of the attested corpus instances will not lead to a conclusive confirmation or rejection (a circumstance typical for non-trivial research questions in the digital humanities). »textklang« therefore supports an extension of the corpus-based methodology by experimental means: By using state-of-the-art speech synthesis techniques, test items for perception experiments can be generated in which the prosodic parameters relevant for the hypothesis can be manipulated at the desired level of granularity, without compromising the naturalness of the sonic realisation. Thus, empirically valid conclusions can be achieved (relative to a suitably formulated version of the initial hypothesis, which of course has to be confined to questions of perception by present-day listeners).

Figure 1 outlines the project's workflow architecture that provides computational support in the researcher's

Estes and Hench (2016), Agirrezabal et al. (2016), Haider and Kuhn (2018)), does address segmental and prosodic aspects of the texts. With few exceptions however (Baumann and Meyer-Sickendiek, 2016; Baumann et al., 2019), corpus-oriented computational work has not started to include recordings of recitations of the poems in the analytical spectrum.

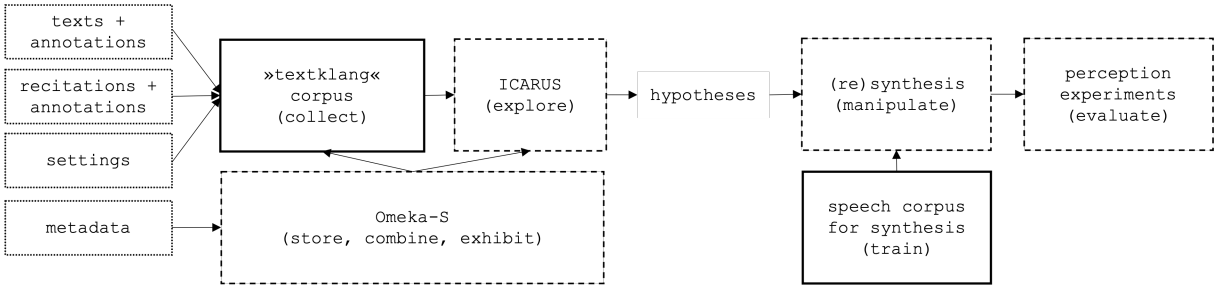

Figure 1: The workflow architecture in the »textklang« project.

research practice. The data (nodes with dotted lines) comprises automatically annotated texts, automatically annotated recitations and, at a later stage, also musical settings, of poetry from the Romantic era. These data constitute the »textklang« corpus (solid line). The boxes with dashed lines represent methodological procedures supported by our infrastructure: Recordings, corresponding texts and metadata for each recording are stored and made available by means of Omeka-S (Omeka, 2022), an open-source repository (see Section 5). The major tool for exploring and visualising the corpus is done by means of ICARUS (Gärtner, M. and Thiele, G., 2021). The hypotheses developed by exploring the data at hand will be evaluated by (re-)synthesising recitations. To this means, we are working on methodologies to adapt neural network based speech synthesis to our specific needs. We aim for a synthesis model that allows to exactly replicate the original recordings and further allows for manipulation of fine-phonetic detail so that hypotheses can be systematically tested in perception experiments. We decided for FastSpeech 2 (Ren et al., 2021) as the underlying model architecture as it inherently offers high controllability. We are implementing our system with the IMS Toucan Speech Synthesis Toolkit (Lux, F., 2022). In order to enhance the training data for the synthesis model with domain specific data, an additional corpus is being created, comprising around 50 hours of recitations of prose and lyric poetry from the same period (i.e. all of Grimm's Fairytales (Grimm and Grimm, 1837) and "Des Knaben Wunderhorn: Alte deutsche Lieder" (*The boy's magic horn: old German songs*), which is a collection of German folk poems and songs edited by Achim von Arnim and Clemens Brentano and first published in three volumes in 1806 and 1808 (Arnim and Brentano, 1808)).

3. Data

The data in our project comprises German poetry from the Romantic era in its various manifestations: The written versions, the recitations of these versions, and, at a later stage, also compositions that have set these poems to music, and the sung performance of these compositions. The following sections focus on the spoken data (Section 3.1) and the corresponding texts (Section 3.2). In addition to the primary data, metadata is collected for each recitation (Section 3.3).

3.1. Audio data

Collection The audio corpus of »textklang« is generated from the collections of the German Literature Archive Marbach which holds a rich collection of recordings of Romantic works, encompassing 2.700 recitations and settings, reaching back to the 1920s and including a broad variety of speakers. Authors whose poetic works will be part of the "textklang" corpus include e.g. Friedrich Hölderlin, Eduard Mörike, Joseph von Eichendorff, Heinrich Heine, Friedrich Schiller, and Johann Wolfgang von Goethe, to name but a few. Analysing spoken poetry from the Romantic era is especially interesting due to the status artists from this era attributed to sound, especially to the musical quality of poetry which became a common topos in the poetological debate of the time (Naumann, 2017; Schneider, 2004). At the same time, especially Romantic poetry is well received from the musical side and set to music.
The recordings available in this collection are stored in various formats, mainly on record, shellac and CD. They are not linked to one respective textual representation in the collection, so that it is not necessarily apparent which edition was recorded.

Digitisation and cutting As a first step in creating our corpus the analogue recordings are digitised and cut into individual tracks such that one track contains the recitation (or musical version) of one poem. In case the recitation is preceded and followed by applause or music, these parts are cut so that only the speech remains. All single-poem recitations are stored as wav-files.

3.2. Textual data

As already mentioned, the exact text sources of the recorded recitations are not known in most of the cases, except the name of the author and sometimes the titles of the individual poems. This is why the first step in creating the text-based corpus is to identify and search for the texts which the speakers recited (or which the composer set to music).

Identify text source In order to facilitate the identification of the respective corresponding text, we use

IMS Speech (Denisov and Vu, 2019) as a first step. IMS Speech is a web-based speech to text tool for German and English speech transcription. Multiple recordings can be uploaded simultaneously by means of an intuitive web-interface and the transcriptions are made available in various formats. We use the transcripts to identify the poem in question and specifically search for it in the online repository TextGrid (TextGrid-Community, 2020), a long-term archive for research data from the humanities, covering world literature from the beginning of the history of printed books to the 20th century and continuing to grow. Most texts are XML/TEI encoded. We store the link to the respective TEI as part of our metadata (see Table 1).

We automatically extract the respective text from the TEI as offered by TextGrid and transform the text into an intermediate format which is processed in our pipeline (see Section 4.1). If the text cannot be found on TextGrid, we resort to the German Text Archive, DTA (BBAW, 2022), which provides a basic corpus of German-language texts across disciplines and genres with a focus from the early 16th to the early 20th century. Individual poems often need to be searched for within larger collections. We save the rendered HTML version offered on DTA.

Manual revision Determining the exact text version of our recitations requires some extensive manual revision. We can use computer-aided speech recognition to assist us in the first step but since we require a perfect one-to-one transcript for the alignment of text and speech (see Section 4.2) and encoding of the poem's form (lines and stanzas), it can only get us so far. It nevertheless provides useful assistance: using only author and title would not be enough to find the text version, since there are often multiple poems sharing the same title. The German poet Friedrich Hölderlin, for example, wrote at least five different poems with the title "Der Winter" (*Winter*).

In addition to different poems sharing the same title, we often find different versions of the same poem with that title. This is on the one hand due to the fact that many poems were not printed during the author's lifetime and there is no edition of last hand, approved by the author, available. Later editions depending on individual editorial choices based on the available manuscripts often vary from one another. Speakers may therefore use different editions and thus different versions of the same poem. One example is, again, the poet Friedrich Hölderlin, and his poem "Die Liebe" (*Love*). In one version, the first stanza consists of a poem which has also been published as an individual poem under the title "Das Unverzeihliche" (*The Unforgivable*), while other versions of this poem do not have this stanza.

In addition to these deviations from some text source the speaker himself may alter the text by, for example, changing the order of some words, omitting syllables, or even swapping lines.

Additionally, speakers may either read the poem's title or start right away with the first line. Some speakers also recite the author's name. So that the text version may be different between different recitations of the very same poem.

The aforementioned reasons make it necessary that, once the text on which the recitation is based was found in one of the online repositories, this text is thoroughly compared to the actual sound recording and manually revised if necessary. The revised texts then build the written part of the corpus. All original sources and changes are documented in the metadata repository so that it is possible to identify and compare different recitations.

3.3. Metadata

Table 1 gives an overview of the metadata we collect for each recording. 'Recording' refers to a recitation or sung performance of one individual poem. Metadata is collected by means of Omeka-S by using a project-specific resource template.

4. Annotations

Once we get a word-for-word transcript of the recording, data preprocessing and the automatic annotation process largely follow the pipeline as developed for the GRAIN corpus (Schweitzer, K. et al., 2018).[2] This includes extensive annotations of both the written material as well as the speech data. The annotation layers are mapped onto each other on the basis of the word token (Schweitzer et al., 2018).

4.1. Automatic annotation of text data

The textual data was automatically tokenised with the TreeTagger (Schmid, H., 2017) and sentences were segmented based on punctuation tokens. Deviating from the original GRAIN pipeline an additional annotation of the endings of verses and stanzas is needed.[3]

Automatic annotations include part of speech tagging by TreeTagger (Schmid, H., 2017) and Stanford Tagger (Stanford NLP Group, 2018) and dependency parses and constituency parses by means of different tools: IMS-SZEGED-CIS (Björkelund et al., 2013), Mate (Bohnet B., 2014), BitPar (Schmid, H., 2004), IMSTrans (Björkelund A., 2015) and Stanford Parser (Stanford NLP Group, 2018).

Combining parses from different tools is particularly advantageous in cases of non-canonical data such as the data at hand. Poetry has a number of characteristics which may be potentially problematic for tools trained on different data (e.g. news papers). Capitalisation at the beginning of each line, for example, may

[2]The landing page of the resource handle lists the involved tools and methods.

[3]Since the GRAIN pipeline annotates the document structure of the edited transcripts of radio interviews, this also had to be adapted to the recitations, e.g. including spoken title and author.

Category	Description
ID	unique identifier
shelf mark	of the recording at the German Literature Archive
author	the person who wrote the poem
title	the poem's title
alternative title	if various titles known
year of publication	year in which the poem was published
interpreter	the person who recites or sings the poem
sex	male or female speaker
speaker profile	a professional speaker, an amateur speaker or a speaker with stage experience
date	year of interpretation
live	recited with or without audience
type	recitation or musical version
composer	in case of musical versions
collection	if poem is part of a collection
edition	on which text edition the recitation is based (if known)
sound carrier	on which medium the original is stored
volume	ID of sound carrier
duration	the recitation's duration
rights	whether the recording is licensed
text link	link to TEI
text source	in which online repository the text was found
read title	whether the speaker reads out the poem's title
read author	whether the author was named in the recitation
closer	whether the recitation is followed by a remark written by the author
changed	whether the recitation deviates from the assumed text source
comment	description of alterations from the text source

Table 1: Metadata categories collected for each recording.

be problematic for PoS tagging. Also syntactic structures and the use of punctuation may deviate from what is typically found in news papers or prose, as can be seen in Example 1, taken from the poem "Der Neckar" (*Neckar*, a river in Southern Germany) by Friedrich Hölderlin.[4]

(1) In deinen Tälern wachte mein Herz mir auf
Zum Leben, deine Wellen umspielten mich,
Und all der holden Hügel, die dich
Wanderer! kennen, ist keiner mir fremd.

Additionally, authors frequently use neologisms, and words are often adjusted due to metrical constraints (e.g. *Allerschütt'rer*, a metrically adjusted neologism used by Hölderlin which describes someone who "unsettles all").

By combining different parsers which themselves are based on different approaches, we can extract confidence estimations for each annotation which show us how many parsers made the same choice. It can be particularly interesting, for example, to investigate instances with low parser agreement, which may indicate problematic cases. Or one may decide to create a subset with only high agreement between the parsers.

The outputs of the dependency parsers are merged into one tree by blending the individual parse trees into one dependency tree while taking the majority vote for each relation into account. (See Schweitzer et al. (2018) for more information on our procedure of merging the parser outputs.)

4.2. Automatic annotations of speech data

The acoustic data is force-aligned for phone, syllable and word boundaries (Rapp S., 2006).

PaIntE parameters are calculated for each syllable in the data (Möhler, 2001; Möhler and Conkie, 1998; Möhler, 1998). PaIntE stands for "Parametrized Intonation Events" and presents a way to model the intonation contour. The model approximates stretches of the F_0-contour by modelling a peak in the smoothed F_0-contour within a three-syllable window by employing a mathematical function term with six free parameters. These parameters depict phonetic cues on and around the accented syllable, namely peak height in Hertz, peak alignment within the three-syllable window, amplitude of the rise, amplitude of the fall, and steepness of the rise and of the fall.

Additional phonetic features are extracted by means of the Festival (Black, A. W. et al., 2004) version of the University of Stuttgart (Festival, 2010) - a synthesis system for German. Among these features are syllable properties such as duration, the position in the word, onset and coda size and type, the number of phonemes in onset and rhyme, the pitch at the beginning, in the middle and at the end of the syllable, the syllable vowel, the syllable vowel duration, whether the syllable is lexically stressed and whether the adjacent syllables are lexically stressed; and word properties such as word duration.

GToBI(S) labels, which are the ToBI annotation labels for German as defined by the Stuttgart system (Mayer, 1995), are automatically annotated as described in Schweitzer (2010). In this procedure, pitch accents and boundary tones are predicted by means of Random Forest classifiers on the bases of the just mentioned PaIntE parameters, as well as normalised phone duration, lexical stress, syllable position, following silence, part of speech and punctuation.

In addition, convolutional neural network (CNN) - based predictions of pitch accent placement are avail-

[4] *In your valleys my heart awakened // To life, your waves played around me, // And of all the lovely hills which you // Wanderer! Know, none is foreign to me.*

able as binary annotations on the word-level. Again, having two different annotation layers for one type of information (such as information on pitch accenting here) can be advantageous when dealing with non-canonical data from a genre for which these models were not trained, by, for example, combining the two approaches to improve the quality of the annotation.
See Schweitzer et al. (2018) for a more detailed presentation of the individual annotation layers.

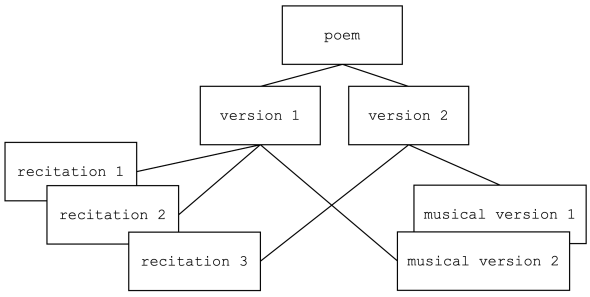

Figure 2: Model depicting the different versions and realisations of one individual poem.

5. Repository

Our repository must meet several requirements: it has to provide the infrastructure for complex data modeling (relations between metadata, textual data and audio data), API for creating, updating and retrieving data, and, additionally, should contain a web publishing platform. Omeka-S (Omeka, 2022), which is an open-source platform for sharing digital collections and creating media-rich online exhibits, satisfies all of our requested features. The great advantage over other repository solutions (e.g., Fedora Commons[5], Dataverse[6]) is, on the one hand, the intuitive maintenance of Omeka-S, which makes it accessible for technically less adept researchers as well, and, on the other hand, the web exhibition function, which provides the opportunity to exhibit data and results on the project website.[7] Additionally, Omeka-S provides different export formats which make the curated data sustainable.
Within Omeka-S, our data are organised such that each recitation is compiled as an individual item with its own metadata and associated text. Recitations (and musical versions) of the same poem can be linked together in one 'item set'. Figure 2 depicts the data model underlying our repository: each recitation or musical version is stored as an item and can be combined hierarchically to larger sets depicting specific versions or poems.

6. Exploration and Visualisation with ICARUS

We use ICARUS (Gärtner, M. and Thiele, G., 2021), a desktop application for multi-modal exploration and visualisation(Gärtner et al., 2015), which makes it possible for the researcher to access and query all annotation layers in our corpus. ICARUS is implemented as a desktop application in Java and features dedicated visualisations for structural aspects of a corpus, such as dependency syntax and coreference, as well as an integration of acoustic information based on PaIntE annotations. It also offers rich customisation options, both for the way in which corpus content is presented, and also regarding the expected physical form of the corpus, i.e. the format to be read: ICARUS supports arbitrary tabular data as input, as long as it is accompanied by a schema definition that tells the tool how to map rows and columns to the internal data model of words, sentences and documents and their annotations, respectively.
This flexibility in terms of input makes it well-suited for use cases such as »textklang« with a very individual corpus structure, especially when it comes to annotations. All annotation layers available for visualisation can also be accessed for querying. Queries in ICARUS follow a simple syntax, with bracketing being used to express structural constraints, and they can be defined either textually or graphically with the integrated query editor. And while ICARUS does not directly support additional resources outside of a corpus for querying, such data can be incorporated into the original corpus format and expressed with a new schema. This way we can adapt the pipeline described in Section 4 to also attach the metadata information shown in Figure 1 to sections in the corpus in order to make it available for researchers to query alongside the raw annotations.

6.1. Example research question: Investigating enjambments

One of the first research questions we pose towards our corpus is how different speakers realise enjambment and what effect the different realisations have on the listener. An enjambment occurs whenever a line ends within a syntactic unit so that versification suggests a prosodic boundary where syntactically the sentence or clause continues.
Speakers can deal with this conflict between syntactic continuation on the one hand and discontinuation as suggested by the line break on the other hand, in different ways. While one speaker may conform to the syntactic continuity and therefore reads over the line, another speaker may follow lineation by realising a prosodic boundary at the end of the line. It has been suggested that speakers may be able to serve both constraints by using cues typically used for the prosodic marking of boundaries and cues typically used within phrases, suggesting continuation, at the same time (Schauffler et al., 2022; Tsur and Gafni, 2019; Tsur, 2012).
The concrete vocal realisation of a text may depend, for example, on the cultural background or the professional education of the speaker, as different schools of recita-

[5] https://github.com/fcrepo4
[6] https://dataverse.org/
[7] http://hdl.handle.net/11022/1007-0000-0007-F6C5-5

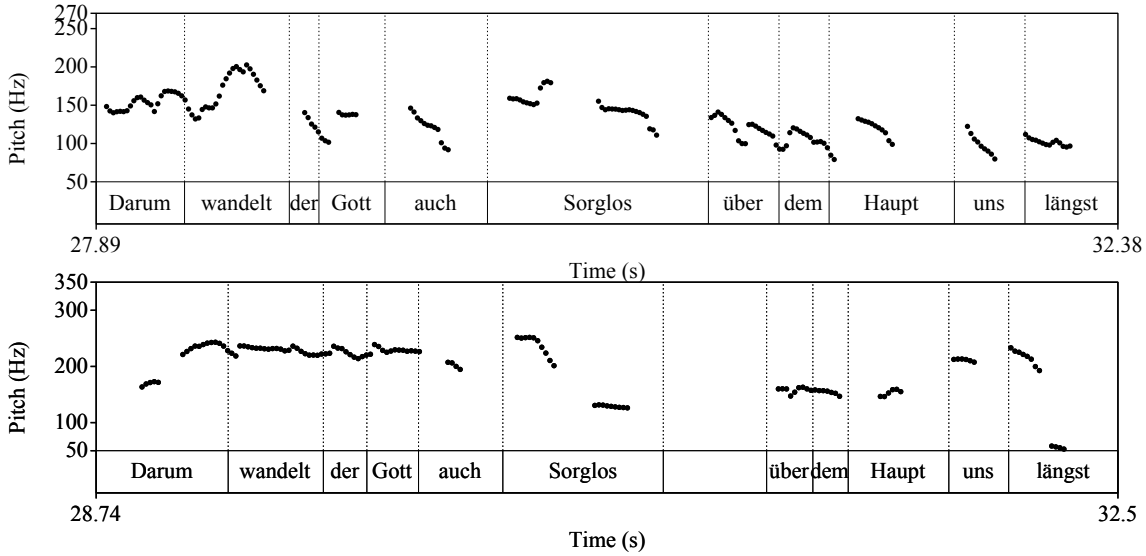

Figure 3: Two versions of an enjambment by two different speakers reciting Hölderlin's poem *Die Liebe*.

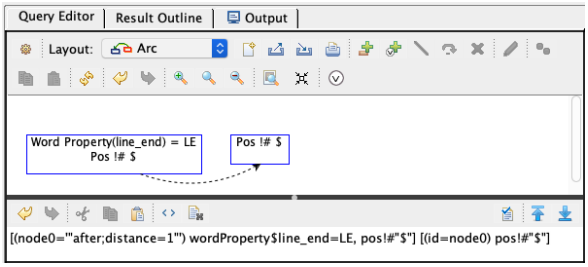

Figure 4: Example search query editor in ICARUS. Top window shows the graphical representation of the search, bottom window shows the search as text.

tion are known to follow particular assumptions and recitation styles (Meyer-Kalkus, 2001; Meyer-Kalkus, 2020).

Figure 3 presents the F_0 contour and speech rate of the two lines presented in Example 2 of Friedrich Hölderlin's poem "Die Liebe" (*Love*) as spoken by two different speakers.[8]

(2) Darum wandelt der Gott auch
Sorglos über dem Haupt uns längst.

The speaker on the top panel realises a phrase break as suggested by versification (visible here as the lowering of F_0 at the end of the line ("auch") and the reset of F_0 at the beginning of the new line ("Sorglos")). The speaker on the bottom panel, on the other hand, reads over the line conforming to syntactic relations, and realises a phrase break only after "Sorglos" (as visible here by the lowering of F_0 after "Sorglos" and the following pause).

ICARUS allows us to systematically search for instances of enjambment in our corpus and to compare different realisations. Figure 4 shows an example query. Since a punctuation mark mostly signals either the completion of a sentence or a clause, we define enjambment as the absence of a punctuation mark (. , ; ! ?) at the end of the line. So in this query we search for tokens which are at the end of lines (and not punctuation marks themselves, left node) and which are not followed by a punctuation mark (right node).

ICARUS provides a comfortable platform to then investigate the output with great detail in the light of all annotations provided on both text and audio. Figure 5 gives an example of how an instance of enjambment can be compared across different speakers. The top window shows details of one instance of enjambment from the query output. The sentence can be played in different granularity, from the whole sentence, over the individual words to the individual syllables. The contours we see here are based on the PaIntE parameters (see Section 4). The annotations presented underneath the syllables can be chosen according to the research interest. All annotations can be investigated. Here, we see the transcribed syllable, the syllable duration and the automatically predicted GToBI(S)-label (see Section 4). In the bottom window, we can see a preview of the following instances, namely realisation of the same lines recited by different speakers. A minimalist intonation preview based on three of the six PaIntE parameters already gives an idea about how different speakers realised the enjambment in terms of intonation.[9]

[8] *Hence God has wandered // With abandon above our head for long.*

[9] Note that duration cues and pauses are not represented in

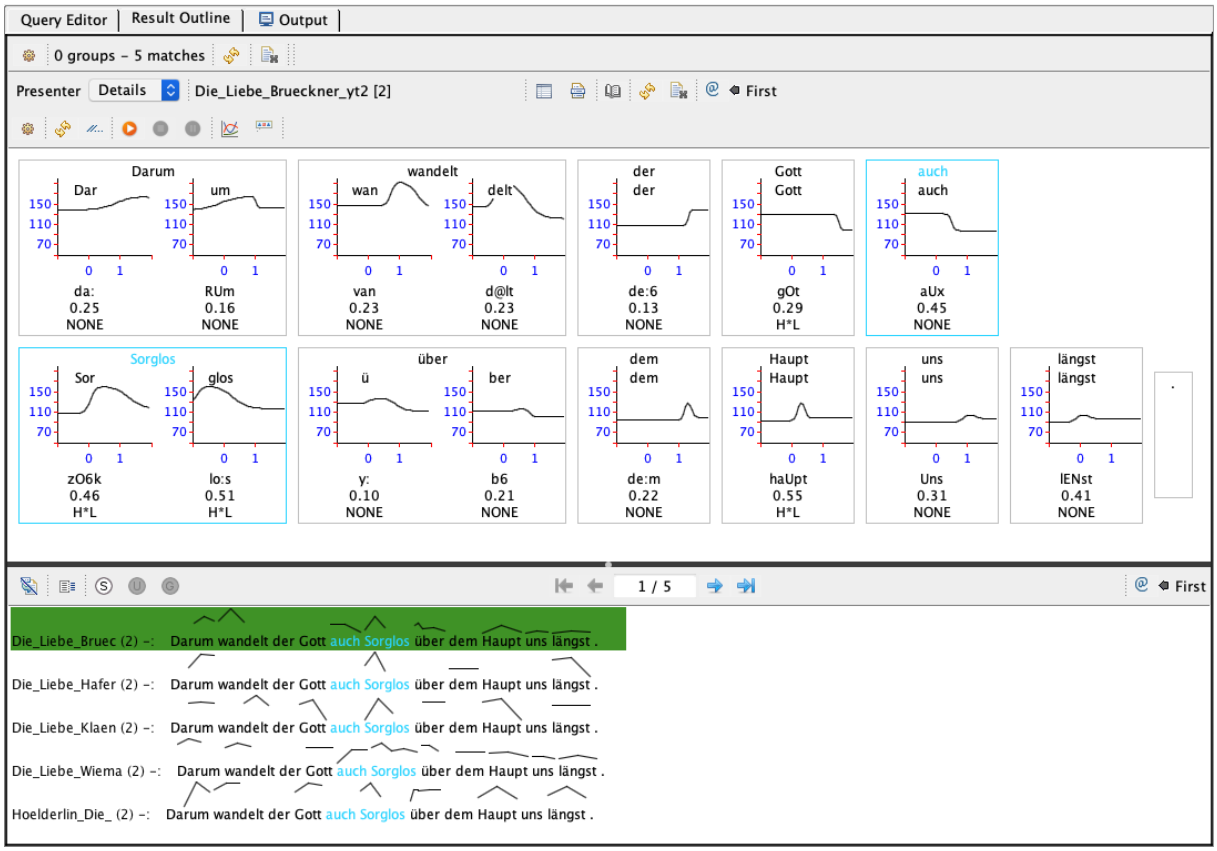

Figure 5: ICARUS presenter platform with detailed information on one instance of the query output on the top window. The bottom window presents a preview of the next instances, here the same lines spoken by different speakers.

7. Conclusion and Outlook

This paper presented the project »textklang« and the steps taken towards a multi-model platform for the investigation of German poetry. We illustrated how in combining established automatic tools with genre-specific fine-tuning and careful manual revision a powerful platform can be created which allows researchers to explore, visualise and evaluate research questions and hypotheses from the humanities.

In our next steps we will increase the toolbox by including automatic annotation of versification, and additionally, we plan to incorporate automatic meter and rhyme detection (Haider, 2021; Haider and Kuhn, 2018). Especially the inclusion of verse information pushes the query capabilities of ICARUS beyond its limits due to the presence of overlapping segmentation layers which the search engine was not designed to handle. To make those additional annotations available for querying we are preparing a web-based faceted search frontend with ICARUS2 (Gärtner M., 2018) as its backend.

With respect to our first research agenda on the realisation of enjambments, we will make use of our multi-dimensional annotation layers to take syntax into account and get a more detailed picture on different kinds of enjambments. The prosodic cues we may identify to be characteristic for a particular style of recitation will be manipulated in (re-)synthesising the original recording. This will give us the opportunity for carefully controlled perception experiments to evaluate the effect the different prosodic choices have on the listener.

An orthogonal perspective of future methodological expansion, which will open up a new interdisciplinary research field, lies in the study of musical settings of Romantic poems. The text-sound relationship can here be addressed both on the basis of the musical scores (which can in part be seen as an approximate access to prosodic realisation strategies in times predating audio recordings) and on the basis of audio recordings of such scores, which of course bring in the singers' choices in vocal interpretation of the score as an additional factor. Examples and developments within the project will be available online.[10]

8. Acknowledgements

This research is supported by funding from the German Ministry for Education and Research (BMBF) for the »textklang« project.

the preview.

[10] http://hdl.handle.net/11022/1007-0000-0007-F6C5-5

9. Bibliographical References

Agirrezabal, M., Alegria, I., and Hulden, M. (2016). Machine learning for metrical analysis of English poetry. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 772–781, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Arnim, A. v. and Brentano, C. (1808). *Des Knaben Wunderhorn. Alte deutsche Lieder. 3 volumes.* Mohr und Zimmer, Heidelberg.

Baumann, T. and Meyer-Sickendiek, B. (2016). Large-scale analysis of spoken free-verse poetry. In *Proceedings of Language Technology Resources and Tools for Digital Humanities (LT4DH)*, Osaka, Japan, dec.

Baumann, T., Hussein, H., Meyer-Sickendiek, B., and Elbeshausen, J. (2019). A tool for human-in-the-loop analysis and exploration of (not only) prosodic classifications for post-modern poetry. In *Proceedings of INF-DH*, pages 151–156, Kassel, Germany, sep. Gesellschaft für Informatik.

Björkelund, A. and Nivre, J. (2015). Non-deterministic oracles for unrestricted non-projective transition-based dependency parsing. In *Proceedings of the 14th International Conference on Parsing Technologies*, pages 76–86, Bilbao, Spain, 07. Association for Computational Linguistics.

Björkelund, A., Çetinoğlu, Ö., Farkas, R., Mueller, T., and Seeker, W. (2013). (Re)ranking meets morphosyntax: State-of-the-art results from the SPMRL 2013 shared task. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 135–145, Seattle, Washington, USA, October. Association for Computational Linguistics.

Black, A. W. (1997). The Festival speech synthesis system. www.cstr.ed.ac.uk/projects/festival.html, November.

Bohnet, B. and Nivre, J. (2012). A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Jeju Island, Korea, 07. Association for Computational Linguistics.

Chen, D. and Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, Doha, Qatar, October. Association for Computational Linguistics.

Denisov, P. and Vu, N. T. (2019). IMS-speech: A speech to text tool. *ESSV*.

Estes, A. and Hench, C. (2016). Supervised machine learning for hybrid meter. In *Proceedings of the Fifth Workshop on Computational Linguistics for Literature*, pages 1–8, San Diego, California, USA, June. Association for Computational Linguistics.

Festival. (2010). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. IMS German Festival home page. `www.ims.uni-stuttgart.de/phonetik/synthesis`.

Gärtner, M., Schweitzer, K., Eckart, K., and Kuhn, J. (2015). Multi-modal visualization and search for text and prosody annotations. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pages 25–30, Beijing, China, July. Association for Computational Linguistics and The Asian Federation of Natural Language Processing.

Gärtner, M. (2020). The corpus query middleware of tomorrow – a proposal for a hybrid corpus query architecture. In *Proceedings of the 8th Workshop on Challenges in the Management of Large Corpora*, pages 31–39, Marseille, France, May. European Language Ressources Association.

Grimm, J. and Grimm, W. (1837). *Kinder und Hausmärchen. Gesammelt durch die Brüder Grimm. 2 volumes*. Dieterich, Göttingen, 3rd edition.

Haider, T. and Kuhn, J. (2018). Supervised rhyme detection with Siamese recurrent networks. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 81–86, Santa Fe, New Mexico, August. Association for Computational Linguistics.

Haider, T. (2021). Metrical tagging in the wild: Building and annotating poetry corpora with rhythmic features. In *16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3715–3725, Kyiv, Ukraine.

Hayes, B., Wilson, B. C. A., and Shisko, B. C. A. (2012). Maxent grammars for the metrics of shakespeare and milton. *Language*, 88:691–731.

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.

Kuhn, J. (2019). Computational text analysis within the humanities: How to combine working practices from the contributing fields? *Language Resources and Evaluation*, 53(4):565–602.

Lux, F., Koch, J., Schweitzer, A., and Vu, N. T. (2021). The IMS Toucan system for the Blizzard Challenge 2021. In *Proc. Blizzard Challenge Workshop*, volume 2021. Speech Synthesis SIG.

Mayer, J. (1995). Transcribing German intonation – the Stuttgart system. Technical report, Universität Stuttgart.

Meyer-Kalkus, R. (2001). *Stimme und Sprechkünste im 20. Jahrhundert*. Akademie Verlag, Berlin.

Meyer-Kalkus, R. (2020). *Geschichte der literarischen Vortragskunst*. J.B. Metzler, Stuttgart.

Möhler, G. and Conkie, A. (1998). Parametric modeling of intonation using vector quantization. In *Proceedings of the Third International Workshop on*

Speech Synthesis (Jenolan Caves, Australia), pages 311–316.

Möhler, G. (1998). Describing intonation with a parametric model. In *Proceedings of the International Conference on Spoken Language Processing*, volume 7, pages 2851–2854.

Möhler, G. (2001). Improvements of the PaIntE model for F_0 parametrization. Technical report, Institute of Natural Language Processing, University of Stuttgart. Draft version.

Naumann, B. (2017). 'Musikalisierung' von Literatur in der Romantik. In Nicola Gess et al., editors, *Handbuch Literatur & Musik*, pages 374–385. De Gruyter, Berlin / Boston.

Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models – An aligner for German. In *Proc. of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry" (Russia)*.

Reiter, N. and Pichler, A. (2020). Reflektierte Textanalyse. In Nils Reiter, et al., editors, *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, pages 43–59. Berlin: De Gruyter.

Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., and Liu, T.-Y. (2021). Fastspeech 2: Fast and high-quality end-to-end text to speech. arXiv:2006.04558, https://doi.org/10.48550/arXiv.2006.04558.

Schauffler, N., Schubö, F., Bernhart, T., Eschenbach, G., Koch, J., Richter, S., Viehhauser, G., Vu, T., Wesemann, L., and Kuhn, J. (2022). Prosodic realisation of enjambment in recitations of German poetry. In *Proc. Speech Prosody*, Lissabon (Portugal).

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK.

Schmid, H. (2004). Efficient parsing of highly ambiguous context-free grammars with bit vectors. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, Geneva, Switzerland.

Schmid, H. (2006). Trace prediction and recovery with unlexicalized pcfgs and slash features. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 177–184, Sydney, Australia, 07. Association for Computational Linguistics.

Schneider, J. N. (2004). *Ins Ohr geschrieben. Lyrik als akustische Kunst zwischen 1750 und 1800*. Wallstein, Göttingen.

Schweitzer, K., Eckart, K., Gärtner, M., Falenska, A., Riester, A., Roesiger, I., Schweitzer, A., Stehwien, S., and Kuhn, J. (2018). German radio interviews: The GRAIN release of the SFB732 silver standard collection. In Nicoletta Calzolari, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 2887–2895, Paris, France, may. European Language Resources Association (ELRA).

Schweitzer, A. (2010). *Production and Perception of Prosodic Events – Evidence from Corpus-based Experiments*. Doctoral dissertation, Universität Stuttgart.

Tsur, R. and Gafni, C. (2019). Enjambment – irony, wit, emotion. A case study suggesting wider principles. *Studia Metrica et Poetica*, 5:7–28, 01.

Tsur, R. (2012). *Poetic Rhythm: Structure and Performance – An Empirical Study in Cognitive Poetics*. Sussex Academic Press, Brighton and Portland.

10. Language Resource References

BBAW. (2022). *Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache.* Accessible via: http://www.deutschestextarchiv.de/.

Björkelund A. (2015). *IMSTrans*. Distributed via: http://hdl.handle.net/11022/1007-0000-0007-F6C7-3, 1.0.

Black, A. W. et al. (2004). *The Festival Speech Synthesis System*. Distributed via: https://www.cstr.ed.ac.uk/projects/festival/, 1.9.5.

Bohnet B. (2014). *Mate Tools*. Distributed via: http://hdl.handle.net/11022/1007-0000-0000-8E4E-A, 3.61.

Denisov, P. and Vu, N. T. (2019). *IMS-Speech*. Distributed via http://hdl.handle.net/11022/1007-0000-0007-F6C6-4.

Gärtner, M. and Thiele, G. (2021). *ICARUS: Interactive platform for Corpus Analysis and Research tools, University of Stuttgart*. Distributed via http://hdl.handle.net/11022/1007-0000-0000-8E56-0, 1.4.9.

Gärtner M. (2018). *ICARUS2: 2nd generation of the Interactive platform for Corpus Analysis and Research tools, University of Stuttgart*. Distributed via: http://hdl.handle.net/11022/1007-0000-0007-C635-E.

Lux, F. (2022). *IMS Speech Synthesis Toolkit Tucan*. Distributed via GitHub: https://github.com/DigitalPhonetics/IMS-Toucan.

Omeka. (2022). *Omeka S*. Omeka, distributed via GitHub: https://github.com/omeka/omeka-s, 3.2.

Rapp S. (2006). *Aligner – an Automatic Speech Segmentation System*. Distributed via: http://hdl.handle.net/11022/1007-0000-0007-F6C8-2, 2.3beta.

Schmid, H. (2004). *BitPar*. Distributed via: http://hdl.handle.net/11022/1007-0000-0000-8E53-3.

Schmid, H. (2017). *TreeTagger*. Distributed via: http://hdl.handle.net/11022/1007-0000-0000-8E4D-B, linux-3.2.1.

Schweitzer, K. et al. (2018). *GRAIN*. Distributed via: http://hdl.handle.net/11022/1007-0000-0007-C631-2, SFB732 Silver Standard Collection, 1.0.

Stanford NLP Group. (2018). *Stanford Parser*. Distributed via: https://nlp.stanford.edu/software/, 3.9.1.

TextGrid-Community. (2020). *TextGrid*. Accessible via: https://textgridrep.org/.