Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 715–722
Marseille, 20-25 June 2022

Querying Interaction Structure: Approaches to Overlap in Spoken Language Corpora

Elena Frick[1], Henrike Helmer[1], Thomas Schmidt[1,2]
[1]Leibniz-Institute for the German Language, R5, 6-13, D-68161 Mannheim, Germany
[2]RISE University of Basel, Spalenberg 65, CH-4051 Basel, Switzerland
{frick, helmer}@ids-mannheim.de, th.schmidt@unibas.ch

Abstract

In this paper, we address two problems in indexing and querying spoken language corpora with overlapping speaker contributions. First, we look into how token distance and token precedence can be measured when multiple primary data streams are available and when transcriptions happen to be tokenized, but are not synchronized with the sound at the level of individual tokens. We propose and experiment with a speaker-based search mode that enables any speaker's transcription tier to be the basic tokenization layer whereby the contributions of other speakers are mapped to this given tier. Secondly, we address two distinct methods of how speaker overlaps can be captured in the TEI-based ISO Standard for Spoken Language Transcriptions (ISO 24624:2016) and how they can be queried by MTAS – an open source Lucene-based search engine for querying text with multilevel annotations. We illustrate the problems, introduce possible solutions and discuss their benefits and drawbacks.

Keywords: spoken language corpora, multi-turn conversations, corpus search engine, query language

1. Introduction

Interaction corpora are collections of audio and/or video recordings of spontaneous and authentic conversations. They differ from corpora of written language and also from some oral corpora (such as phonetic corpora) because they contain verbal interactions between two or more interlocutors and therefore have multiple primary data streams. Methodological and technical challenges for working with this special type of corpus are described in Schmidt (2018). In the present paper, we focus on two specific problems arising when indexing and searching interaction corpora. In particular, we look first into how token distance and token precedence can be measured in spoken language transcripts with overlapping speaker contributions containing tokens that are not synchronized with the audio sound. Secondly, we address two distinct methods of how speaker overlaps can be algorithmically computed and stored in the TEI-based ISO Standard for Spoken Language Transcriptions (ISO 24624:2016).

The paper is organized as follows: Section 2 briefly explains the background and motivation of our study. Section 3 presents our methods in indexing and searching interaction corpora and proposes some possible solutions in dealing with speaker overlaps. The paper continues with related work in Section 4 and provides the conclusion of our research in Section 5.

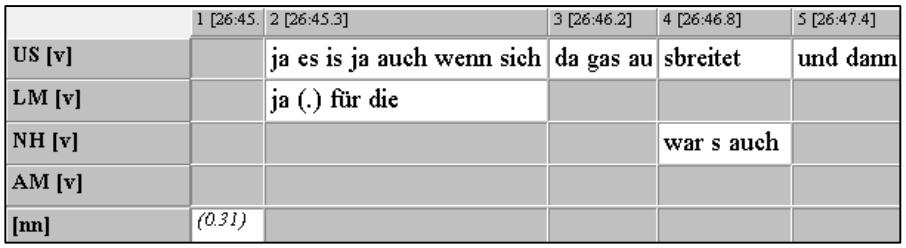
2. Background and Motivation

The background of the present study is the project ZuMult (Zugänge zu multimodalen Korpora gesprochener Sprache, Access to Multimodal Spoken Language Corpora)[1]. It is a DFG-funded three-year cooperation project between the Archive of Spoken German (AGD)[2] in Mannheim, the Hamburg Centre for Language Corpora (HZSK)[3] and the Herder-Institute[4] at the University of Leipzig. One of the aims of ZuMult is to develop a backend software architecture for a unified access to spoken language resources located in different repositories (cf. Batinić et al., 2019; Fandrych et al., 2022). The access should also include the search functionality allowing to query corpora stored in the TEI-based ISO Standard for Spoken Language Transcripts. For this purpose, we explored how MTAS (Brouwer, Brugman, and Kemps-Snijders 2016) – an open source Lucene-based search engine framework developed for querying texts with multilevel annotations – can be reused for searching spoken language corpora. The corpora we are dealing with are interaction corpora for the most part (cf. e.g. FOLK[5], GeWiss[6], HaMaTaC[7]). We were interested whether this special type of corpora can be indexed with MTAS and searched by using its query language, a modified version of the CQP Query Language originally developed for the IMS Open Corpus Workbench (CWB)[8]. We introduced the first results of this research in Frick and Schmidt (2020) where we outlined the capacity, but also the limitations of MTAS in terms of its compatibility with typical characteristics of spoken language. The present paper continues this work and addresses two challenging issues concerning speaker overlaps in corpora without complete token-based time-alignment.

3. Methods

In this section, we illustrate the problems every corpus research tool developer has faced sooner or later when implementing search software for interaction corpora. The first problem concerns the token distance and token precedence within speaker overlaps. The other one relates to the possibilities for indexing and searching speaker overlaps. We propose some solutions implemented with MTAS and discuss their benefits and disadvantages.

[1] https://zumult.org/
[2] http://agd.ids-mannheim.de
[3] https://corpora.uni-hamburg.de/hzsk/
[4] https://www.philol.uni-leipzig.de/herder-institut/
[5] http://agd.ids-mannheim.de/folk.shtml
[6] https://gewiss.uni-leipzig.de
[7] https://corpora.uni-hamburg.de/hzsk/de/islandora/object/spoken-corpus:hamatac
[8] http://cwb.sourceforge.net/



	1 [26:45.	2 [26:45.3]	3 [26:46.2]	4 [26:46.8]	5 [26:47.4]
US [v]		ja es is ja auch wenn sich	da gas au	sbreitet	und dann
LM [v]		ja (.) für die			
NH [v]				war s auch	
AM [v]					
[nn]	(0.31)				

Figure 1: An excerpt of the FOLK corpus transcriptions (FOLK_E_00055_SE_01_T_03) opened in EXMARaLDA.

3.1 Token Distance and Precedence

3.1.1 Problem

Compared to written corpora, indexing and querying the token distance in spoken language transcriptions is not a straightforward task, because it is not clearly determined what elements of a transcript (word tokens, transcribed pauses, non-verbal sounds, time anchors etc.) should be considered as an equivalent to a text token (see Frick and Schmidt, 2020). But even if this question has been clarified, multiple speaker layers with overlapping contributions are a particular problem for dealing with token distance and token precedence in interaction corpora.

In spoken language transcripts, we generally have to deal with two token orders: temporal token order and document/sequential order of tokens, which don't coincide in the case of speaker overlaps. As an example, compare the representations of a transcript excerpt from the FOLK corpus, shown in the EXMARaLDA[9] editor (Figure 1[10]) and in the XML document corresponding to the ISO-TEI Standard for Spoken Language Transcriptions (Figure 2[11]). As you can see, the speakers US and LM are speaking simultaneously for a second, both start their contributions with the word-token *ja* (Eng. *yes*). In the temporal token order, illustrated by the representation in EXMARaLDA, both *ja* are preceded by a pause of 0.31 seconds. In the XML document, the parallel speaker contributions are presented in the sequential order, with the longest contribution (represented in the ISO/TEI standard by the <annotationBlock>-element) occurring before the shorter one. Therefore, only *ja* realized by speaker US is preceded by a pause. The *ja* of speaker LM occurs after the whole contribution of speaker US and follows the token *dann* (Eng. *then*). Furthermore, although the word tokens *ja* of both speakers overlap, the token distance between these words according to the transcript would be 12, because 11 tokens occur between them in the XML file (see <w>-elements with xml:id w3007 and w3019 in Figure 2, marked by boxes).

Because of efficiency in transcribing, the audio alignment is usually made in units above the word level (e.g. utterance units or longer contributions) and many individual tokens in the transcripts are therefore not synchronized with the audio sound. In theory, a word (or even phoneme) level alignment could be added with forced aligners such as MAUS[12]. In practice however, such an alignment would be highly unreliable especially in the overlapping passages because forced aligners have no way of dealing with simultaneous speech (making multi-channel recordings is usually not a viable option for this type of field recording). So, in this case, the temporal token order cannot be determined anymore and only the document order can be used to measure the distance and precedence of tokens. This leads sometimes to incorrect, incomplete or misleading results when searching token sequences. For example, the following CQP query looks for all interjections and response particles (POS-tag: NGIRR) realized after a pause, but the search engine working on the document token order will match only *ja* of speaker US and not the other occurrence of *ja* realized by speaker LM in the transcription excerpt discussed above.

[pos="NGIRR"] precededby <pause/>

looks for tokens that are annotated with 'NGIRR' at the POS layer and follow a pause

3.1.2 Solution

Trying to solve the problems described in the section before, we tested the so-called *Speaker-based search mode*. In this approach we created the speaker-based versions of each transcript, which means that every speaker of the transcript got a separate document containing only the transcriptions of this speaker. All annotations and transcriptions of other speakers and speakerless elements (such as pauses between contributions) were mapped to this new tokenization layer. After indexing these speaker-based documents with the MTAS-based search engine, we could search in our corpora by individual speakers. That means that the query string from Section 3.1.1 can now match both occurrences of *ja* in the example presented in Figures 1 and 2.

Furthermore, in the speaker-based transcripts, we could automatically add new time-based span annotations marking all time intervals when the speaker is silent, but other speakers are speaking. For example, the following <spanGrp>-element was added to the speaker-based ISO-TEI transcript corresponding to speaker NH from the example in Figures 1 und 2.

[9] https://exmaralda.org/de/

[10] In order to save space, we are not providing an English translation for the German material. We trust that this will not keep the reader from following our arguments, which are about the structural properties, not the meaning of tokens.

[11] Please note that, for the sake of readability, we have simplified the XML to only display the information needed to understand the argumentation in this paper. The full ISO compliant version contains additional attributes on most elements, most importantly normalisation (@norm), lemma (@lemma) and pos (@pos) annotation for each token.

[12] https://www.bas.uni-muenchen.de/Bas/BasMAUS.html

```
<pause dur="PT0.31S" end="TLI_992" rend="(0.31)" start="TLI_991" xml:id="p301"/>
<annotationBlock xml:id="c667" end="TLI_996" start="TLI_992" who="US">
   <u>
    <seg>
      <anchor synch="TLI_992" type="ol-start"/>
      <w xml:id="w3007" type="ol-in" lemma="ja" norm="ja" pos="NGIRR">ja</w>
      <w xml:id="w3008" type="ol-in">es</w>
      <w xml:id="w3009" type="ol-in">is</w>
      <w xml:id="w3010" type="ol-in" >ja</w>
      <w xml:id="w3011" type="ol-in" >auch</w>
      <w xml:id="w3012" type="ol-in" >wenn</w>
      <w xml:id="w3013" type="ol-in" >sich</w>
      <anchor synch="TLI_993" type="ol-end"/>
      <w xml:id="w3014">da</w>
      <w xml:id="w3015">gas</w>
      <w xml:id="w3016" type="ol-in">au<anchor synch="TLI_994" type="ol-start"/>sbreitet</w>
      <anchor synch="TLI_995" type="ol-in"/>
      <w xml:id="w3017" type="ol-in" >und</w>
      <w xml:id="w3018" type="ol-in">dann</w>
      <anchor synch="TLI_996" type="ol-end"/>
    </seg>
   </u>
   <spanGrp type="speaker-overlap" subtype="time-based">
      <span from="TLI_992" to="TLI_993">LM</span>
      <span from="TLI_994" to="TLI_995">NH</span>
      <span from="TLI_995" to="TLI_996">AM</span>
   </spanGrp>
</annotationBlock>
<annotationBlock end="TLI_993" start="TLI_992" who="LM" xml:id="c668">
   <u>
    <seg>
      <anchor synch="TLI_992" type="ol-start"/>
      <w xml:id="w3019" type="ol-in" lemma="ja" norm="ja" pos="NGIRR">ja</w>
      <pause rend="(.)" type="micro" xml:id="p302"/>
      <w xml:id="w3020" type="ol-in">für</w>
      <w xml:id="w3021" type="ol-in">die</w>
      <anchor synch="TLI_993" type="ol-end"/>
    </seg>
   </u>
   <spanGrp type="speaker-overlap" subtype="time-based">
      <span from="TLI_992" to="TLI_993">US</span>
   </spanGrp>
</annotationBlock>
<annotationBlock xml:id="c669" end="TLI_995" start="TLI_994" who="NH">
  <u>
   <seg>
     <anchor synch="TLI_994" type="ol-start"/>
     <w xml:id="w3022" type="ol-in">war</w>
     <w xml:id="w3023" type="assimilated ol-in" >s</w>
     <w xml:id="w3024" type="ol-in" >auch</w>
     <anchor synch="TLI_995" type="ol-end"/>
   </seg>
  </u>
  <spanGrp type="speaker-overlap" subtype="time-based">
     <span from="TLI_994" to="TLI_995">US</span>
  </spanGrp>
</annotationBlock>
```

Figure 2: The same excerpt of the FOLK corpus as in Figure 1 presented in the ISO-TEI standard.

```
<spanGrp type="another-speaker" subtype="time-based">
   <span from="TLI_992" to="TLI_994">US</span>
   <span from="TLI_992" to="TLI_993">LM</span>
   <span from="TLI_995" to="TLI_996">US</span>
<spanGrp/>
```

Using these annotations, users can now perform complex searches by taking into account phenomena like speaker-change and turn-taking as demonstrated in the queries below.

([norm="oder"] !within <speaker-overlap/>) followedby <para/>{0,5}<another-speaker/>

looks for any transcribed form of 'oder' occurring outside of an overlap at the last position before speaker change; 'para' stands for <pause>-, <vocal>- and <incident>-elements which can occur in the transcription between two speaker contributions.

	9 [51:57.5]	10 [51:58.	11 [51:58.5]	12 [51:58.9]	13 [52:00.2]	14 [52:00.5]	15 [52:01.4]
AM [v]			((lacht))			ja des nein nein	es is
US [v]	ja so (.)	nee de	s klingt d	es klingt jetzt vielleicht hart ich weiß	nich wie ich s	besser ausdrücken	sol
NH [v]		((lacht)) das jetz vielleicht über			trieben		ja
[nn]							

Figure 3: A transcript excerpt (FOLK_E_00055_SE_01_T_05) demonstrating the problem for the segment-based approach proposed in Section 3.2.2.

(<annotationBlock/> containing ([word=".*" & !pos="(NGIRR|NGHES|XY)"] !within <speaker-overlap/>)) precededby (<another-speaker/><para/>{0,5})

looks for turn-taking by one of the non-speakers whose contribution contains at least one word token that occurs outside of the speaker overlap and is not a non-word, hesitation, interjection or responsive particle.

3.1.3 Discussion

The speaker-based search mode does not make the common transcript-based search superfluous, but it complements its search options in a very useful way as shown by the search examples above.
However, this additional search approach comes at a price: a lot of storage space for additional search indices is required, and the computational time needed for corpus indexing increases strongly depending on the number of speakers (consider classroom interactions with dozens of students).

3.2 Speaker Overlaps

3.2.1 Problem

The search functionality developed in the ZuMult project was designed with special user groups in mind. In particular, these are conversation analysis researchers interested in a new corpus search environment that makes it possible, among other things, to search for features of interaction structure, such as speaker overlaps.
Although the MTAS framework used in the ZuMult search engine supports the search for overlapping structures and annotations, the MTAS Query Language is limited on the level of syntax to allow flexible searches for speaker overlaps. For example, it is possible to use the MTAS Query Language operator “intersecting” to search for contributions of speaker A overlapping with contributions of speaker B:

<annotationBlock.speaker="A"/>
intersecting <annotationBlock.speaker="B"/>

But, it is not possible to write a query looking for all speaker overlaps in general. The query expression like

<annotationBlock.speaker/> intersecting
<annotationBlock.speaker/>

would match every speaker's contribution because it would overlap with itself. To get the desired search result, the query should be formulated in a way like this:

<annotationBlock.speaker=$X/> intersecting
<annotationBlock.speaker=$Y/> where $X!=$Y

However, this form of using variables is not supported in the current version of the MTAS Query Language.

3.2.2 Solution

Extending the MTAS Query Language syntax to support variables is not a practicable option for us, because we use MTAS as an embedded framework that is being developed outside of our project. We did not want to change the framework itself in order to remain flexible and to be able to switch easily to the newest version of MTAS at any time later.
The solution we chose to allow users to search for speaker overlaps was adding the appropriate annotations to the transcript documents and storing them in the MTAS search index. The ISO-TEI structure and the content of our transcripts allow for different methods to automatically identify speaker overlaps. The annotations of speaker overlaps can also be added in various forms to the transcript document. Consequently, we decided to test two different methods by adding two different kinds of annotations and to compare them to validate their effectiveness.
The first method is segment-based. It goes through the time segments in the tokenization layer and checks for each pair of time anchors whether there are equivalents in the contributions of other speakers. If time anchors with the same value in the *synch*-attribute could be found in the contribution of another speaker, they are marked as the start and the end of a speaker overlap (see e.g. the *type*-attribute of the <anchor>-elements containing the *synch*-attribute with values TLI_992 and TLI_993 in Figure 2, marked by grey highlighting) and all word tokens between them get an annotation tag “ol-in” (“within overlap”) in the *type*-attribute (see e.g. <w>-elements with xml:id w3007-w3013 in Figure 2).
The *type*-attribute was indexed using MTAS in the same way as other token-based annotations like transcribed and normalized forms, POS-tags and lemmas. This allowed the following types of search queries to be submitted over the ZuMult Search-API:

[word.type=".*ol-in.*"]
looks for word tokens within overlaps; the search pattern containing regular expression characters ‘.’ from both sides of ‘ol-in’ is important to match also type-attributes containing multi-word values (see e.g. the type-attribute of w3023 in Figure 2)*

```
<annotationBlock xml:id="c150" start="TLI_249" end="TLI_257" who="US">
  <u>
    <seg>
      <anchor synch="TLI_249"/>
      ...
      <anchor synch="TLI_250" type="ol-start"/>
      <w xml:id="w854" type="ol-in">nee</w>
      <w xml:id="w855" type="ol-in">de<anchor synch="TLI_251"/>s</w>
      <w xml:id="w856" type="ol-in">klingt</w>
      <w xml:id="w857" type="ol-in">d<anchor synch="TLI_252" type="ol-end"/>es</w>
      <w xml:id="w858">klingt</w>
      <w xml:id="w859">jetzt</w>
      <w xml:id="w860">vielleicht</w>
      <w xml:id="w861">hart</w>
      <w xml:id="w862">ich</w>
      <w xml:id="w863">weiß</w>
      <anchor synch="TLI_253" type="ol-start"/>
      <w xml:id="w864" type="ol-in">nich</w>
      <w xml:id="w865" type="ol-in">wie</w>
      <w xml:id="w866" type="ol-in">ich</w>
      <w xml:id="w867" type="assimilated ol-in">s</w>
      <anchor synch="TLI_254" .../>
      ...
    </seg>
  </u>
  <spanGrp type="speaker-overlap" subtype="time-based">
    <span from="TLI_250" to="TLI_254">NH</span>
    <span from="TLI_251" to="TLI_252">AM</span>
    <span from="TLI_254" to="TLI_257">AM</span>
    <span from="TLI_255" to="TLI_256">NH</span>
  </spanGrp>
</annotationBlock>
<annotationBlock xml:id="c151" start="TLI_250" end="TLI_254" who="NH">
  <u>
    <seg>
      <anchor synch="TLI_250"/>
      <incident>
        <desc rend="((lacht))">lacht</desc>
      </incident>
      <w xml:id="w871">das</w>
      <w xml:id="w872">jetz</w>
      <w xml:id="w873">vielleicht</w>
      <w xml:id="w874" type="ol-in">über<anchor synch="TLI_253" type="ol-start"/>trieben</w>
      <anchor synch="TLI_254" type="ol-end"/>
    </seg>
  </u>
  <spanGrp type="speaker-overlap" subtype="time-based">
    <span from="TLI_250" to="TLI_254">US</span>
    <span from="TLI_251" to="TLI_252">AM</span>
  </spanGrp>
</annotationBlock>
```

Figure 4: The same excerpt of the FOLK corpus as in Figure 3 presented in the ISO-TEI standard.

[norm="bitte" & word.type=".*ol-in.*"]
looks for any transcribed form of 'bitte' within overlaps

<annotationBlock/> containing [word.type=".*ol-in.*"]
looks for all speaker contributions containing overlaps

The second method is contribution-based. It compares the start and end times of each <annotationBlock>-element with the start and end times of all other <annotationBlock>-elements containing contributions of other speakers. If overlaps are identified, the <spanGrp>-element with the start and end times of the overlapping token sequence is added to the <annotationBlock> (see <spanGrp>-elements in Figure 2). The following query expressions demonstrate how the added span annotations can be requested when searching for speaker overlaps:

<speaker-overlap/>
looks for all spans annotated as speaker overlap

<speaker-overlap/> containing [lemma="(Herr|Frau)"]
looks for all spans annotated as speaker overlap and containing any forms of 'Herr' or 'Frau'

<speaker-overlap>[norm="also"]
looks for any transcribed form of 'also' at the beginning of speaker overlaps

<speaker-overlap="SZ"/>
looks for all token sequences overlapping with the contributions of the speaker 'SZ'

	1 [52:41.5]	2 [52:43.	3 [52:43.9]	4 [52:44.4]	5 [52:45.
US [v]	((Lachansatz)) (.) der muss schal schmecken (.) ((Lachansatz))	°h ((lacht))		also trinken ((lacht))	
NH [v]	also der muss (.) eigentlich muss er weg	°h ((lacht))			
AM [v]			war der	war der im kühlschrank also is (.) is der	kalt
[nn]					

Figure 5: A transcript excerpt (FOLK_E_00055_SE_01_T_05) demonstrating the problem for the contribution-based approach proposed in Section 3.2.2

3.2.3 Discussion

Our experimental work showed that none of these methods can be used to index and search ALL speaker overlaps occurring in our corpora. The reason for this is trivial. Some time anchors that would be required for calculating and for indexing overlaps are missing. Please have a look at the example given in the EXMARaLDA editor in Figure 3. In this excerpt from the FOLK corpus, speakers US and NH are speaking simultaneously. If we look at the ISO-TEI representation of the same excerpt in Figure 4, we discover that the time anchor TLI_252 occurring in the speech of US is missing in the contribution of speaker NH. This is because the simultaneity of three speaker contributions makes it impossible in this case for the transcriber to precisely determine where each overlap starts or ends in relation to each of the other contributions. Therefore, the segment-based method could not recognize the word tokens with xml:id w858-w863 as being within the speaker overlap. The contribution-based method is in this case more accurate because it detects the speaker overlaps by comparing the end and start times of the <annotationBlock>-element (see <annotationBlock> with xml:id c150 and the first span annotation entry of its <spanGrp>-element).

However, the contribution-based approach also has its disadvantages. Although it produces the correct time annotations, these annotations could not always be mapped to the tokenization layer during the indexing process, because relevant time anchors are again missing within <annotationBlock>-elements. This is illustrated by the FOLK excerpt in Figure 5, where speaker AM starts talking while speaker US is laughing. For a while they are speaking simultaneously. Using the contribution-based method, the interval with the appropriate speaker overlap can be determined and annotated in the transcript (see Figure 6). Unfortunately, the MTAS indexing algorithm fails when mapping the span annotation to the transcription layer because the time anchor with the *synch*-attribute value T_321 cannot be found in the <annotationBlock>-element of speaker US. The span annotation is simply left out of the search index. That means, that the following query will not match the tokens ‘also’ (w1076) and ‘trinken’ (w1077) in the current example.

<word/> within <speaker-overlap/>
looks for word tokens annotated as speaker overlap

Nevertheless, these both tokens can be found by searching ‘ol-in’ as value of the *type*-attribute as it is shown in the first query example from Section 3.2.2.

Since both methods discussed here have their drawbacks, we propose to use them complementary to each other to get an optimal set of results. Here is an example of a query expression combining both techniques for searching words within speaker overlaps:

(<word/> within <speaker-overlap/> | [word.type=".*ol-in.*"])
looks for word tokens occurring within speaker overlaps

There remains an open question, however, how successful the combination of both methods is. To be able to answer this question, we need manual annotations of speaker overlaps against which the search query below could be evaluated.

We are aware that adding annotations to the transcript documents has disadvantages compared to adapting the MTAS Query Language, mainly because additional storage capacity is required. But our work allows us to conclude that just adding variables to the MTAS Query Language syntax and combining them with the “intersecting” operator (as previously suspected) will not return ALL speaker overlaps occurring in the corpus. A combination of different algorithms for calculating speaker overlaps behind the “intersecting”-operator would be required.

4. Related Work

At the beginning of the ZuMult-project, we have gained an overview of freely available web applications providing online access to spoken language corpora (cf. Batinić, Frick and Schmidt, 2021). Many of these search platforms support the search functionality allowing the token distance specification between the items of the desired word-token sequence (cf. e.g. CQPWeb[13]/BNC2014, Kontext[14], TalkBankDB[15], GLOSSA[16]). But they only take into account the sequential word token order in the document without considering problems caused by speaker overlaps. Support for querying tokens in relation to overlaps is provided by CLAPI[17]. Moreover, this corpus search platform works, among others, with TEI-based transcript format like in our approach. Nevertheless, the CLAPI search possibilities are restricted: it allows for example to search for word tokens followed or preceded by overlaps, but not located within or outside overlaps. In contrast, the Database for Spoken German (Datenbank für Gesprochenes Deutsch, DGD)[18], has a "position filter", which can, for corpora with the respective information encoded, searches to positions within and outside overlaps,

[13] https://cqpweb.lancs.ac.uk/bnc2014spoken/
[14] https://www.korpus.cz/
[15] https://talkbank.org/
[16] https://tekstlab.uio.no/glossa2
[17] http://clapi.ish-lyon.cnrs.fr
[18] https://dgd.ids-mannheim.de

```
<annotationBlock start="TLI_319" end="TLI_323" who="US" xml:id="c187">
  <u xml:id="u_d437e4492">

    <seg type="contribution" xml:id="seg_d437e4492">
      <anchor synch="TLI_319" type="ol-start"/>
      <incident xml:id="n34">
        <desc rend="((Lachansatz))">Lachansatz</desc>
      </incident>
      <pause rend="(.)" type="micro" xml:id="p69"/>
      <w lemma="die" norm="der" pos="PDS" type="ol-in" xml:id="w1072">der</w>
      <w lemma="müssen" norm="muss" pos="VMFIN" type="ol-in" xml:id="w1073">muss</w>
      <w lemma="schal" norm="schal" pos="ADJD" type="ol-in" xml:id="w1074">schal</w>
      <w lemma="schmecken" norm="schmecken" pos="VVFIN"
         type="ol-in" xml:id="w1075">schmecken</w>
      <pause rend="(.)" type="micro" xml:id="p70"/>
      <incident xml:id="n35">
        <desc rend="((Lachansatz))">Lachansatz</desc>
      </incident>
      <anchor synch="TLI_320" type="ol-in"/>
      <vocal xml:id="b30">
        <desc rend="°h">short breathe in</desc>
      </vocal>
      <incident xml:id="n36">
        <desc rend="((lacht))">lacht</desc>
      </incident>
      <anchor synch="TLI_322" type="ol-in"/>
      <w lemma="also" norm="also" pos="SEDM" type="ol-in" xml:id="w1076">also</w>
      <w lemma="trinken" norm="trinken" pos="VVINF"
         type="uncertain ol-in" xml:id="w1077">trinken</w>
      <incident xml:id="n37">
        <desc rend="((lacht))">lacht</desc>
      </incident>
      <anchor synch="TLI_323" type="ol-end"/>
    </seg>
  </u>
  <spanGrp type="speaker-overlap" subtype="time-based">
    <span from="TLI_319" to="TLI_322">NH</span>
    <span from="TLI_321" to="TLI_323">AM</span>
  </spanGrp>
</annotationBlock>
```

Figure 6: The same excerpt of the FOLK corpus as in Figure 5 presented in the ISO-TEI standard.

but DGD again does not support querying and displaying speaker overlaps containing specified word tokens or word token sequences. Both, CLAPI and DGD use the query builder with a complex filter to specify the distance between individual tokens. Compared to CLAPI, DGD provides additionally a speaker-based search mode comparable to the one described here, but the DGD's data model for transcripts is not TEI-based and supports only a fixed set on tokens, no free span annotations. The MTAS-based search engine developed in the ZuMult-project as well as our first prototypical user interface application *ZuRecht*[19] combine both approaches and complement them by using a query language with CQP-based syntax for querying various aspects of speaker overlaps in the ISO-TEI transcript format.

5. Conclusion

The aim of this paper was to draw attention to the difficulties encountered in the development of query systems for interaction corpora. Using two specific phenomena (token distance and speaker overlaps), we have shown how complex such corpora are, especially if they lack the word-token-based time-alignment.

From our point of view, the proposed MTAS-based solutions are helpful to satisfy most of the needs of end users searching in this specific type of corpora[20]. But the optimal answer to the described problems is and remains the time-alignment at the token level. It would allow more precise searches corresponding to token distance and speaker overlaps.

As long as it is not possible to build on the token-based time-alignment, the alternative solutions are welcome and important to be shared with the research community. With the present paper we intend to motivate for more transparency and exchange in the development of the corpus search software for spoken language corpora. As an outlook, we think that the present paper can also provide some discussion material for modelling *Use Cases* in the

[19] http://zumult.ids-mannheim.de/ProtoZumult/jsp/zuRecht.jsp

[20] In April 2022, almost 15000 (inter-)national users are registered for the DGD and thus are potential users of ZuRecht. The fact that our corpora are actively used is proved by numerous publications based on FOLK – the main corpus provided by DGD and ZuRecht. A website collecting these publications is available at www.ids-mannheim.de/prag/muendlichekorpora/bibliographie-folk

"CQLF Ontology for Multi-Stream Architectures" – Part 3 of Corpus Query Lingua Franca (CQLF, ISO 24623-1:2018, for more information about CQLF see Bański, Frick and Witt (2016) and Evert et al. (2020)).

6. Bibliographical References

Bański, P., Frick, E., and Witt, A. (2016). Corpus Query Lingua Franca (CQLF). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2804–2809, Portorož, Slovenia, May 23-28. European Language Resource Association (ELRA).

Batinić, J.,Frick, E., Gasch, J. and Schmidt, T.(2019). Eine Basis-Architektur für den Zugriff auf multimodale Korpora gesprochener Sprache. In Sahle, P. (Ed.), *Digital Humanities: multimedial & multimodal. Konferenzabstracts zur 6. Tagung des Verbandes Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019)*. Frankfurt/Main; Mainz: Verband Digital Humanities im deutschsprachigen Raum e.V., pp. 280–281.

Batinić, J., Frick, E., and Schmidt, T. (2021). Accessing spoken language corpora: an overview of current approaches. In *Corpora*, 16 (3):417–445. Edinburgh: Edinburgh University Press.

Brouwer, M., Brugman, H., and Kemps-Snijders, M. (2016). MTAS: A Solr/Lucene based Multi-Tier Annotation Search solution. *Selected papers from the CLARIN Annual Conference 2016*. Aix-en-Provence, pp. 19–37.

Evert, S., Harlamov, O., Heinrich, P., and Bański, P. (2020). Corpus Query Lingua Franca part II: Ontology. In *Proceedings of the* Twelfth *International Conference on Language Resources and Evaluation (LREC 2020*), pages 3346–3352, Paris: European Language Resources Association (ELRA).

Fandrych, C., Frick, E., Kaiser, J., Meißner, C., Portmann, A., Schmidt, T., Schwendemann, M., Wallner, F., and Wörner, K. (in print). ZuMult: Neue Zugangswege zu Korpora gesprochener Sprache. In Kämper, H. et al. (Eds.), *Sprache in Politik und Gesellschaft: Perspektiven und Zugänge*. Jahrbuch des Instituts für Deutsche Sprache 2021. Berlin etc.: de Gruyter.

Frick, E. and Schmidt, T. (2020). Using Full Text Indices for Querying Spoken Language Data. In *Proceedings of the LREC 2020 Workshop, Language Resources and Evaluation Conference, 11–16 May 2020, 8th Workshop on Challenges in the Management of Large Corpora (CMLC-8)*, pages 40–46, Paris: European Language Resources Association (ELRA).

ISO 24624:2016. Language resource management — Transcription of spoken language.

ISO 24623-1:2018. Language resource management — Corpus query lingua franca (CQLF) — Part 1: Metamodel.

Schmidt, T. (2018). Gesprächskorpora. In M. Kupietz & T. Schmidt (Eds.), *Korpuslinguistik*. (=Germanistische Sprachwissenschaft um 2020, Bd. 5). Berlin/Boston: de Gruyter, pp. 209–230.