Generating a Yiddish Speech Corpus, Forced Aligner and Basic ASR System for the AHEYM Project

Małgorzata E. Ćavar, Damir Cavar, Dov-Ber Kerler, Anya Quilitsch
Indiana University
Bloomington, IN, USA
{mcavar,dcavar,dkerler,aquilitz}@indiana.edu

Abstract

To create automatic transcription and annotation tools for the AHEYM corpus of recorded interviews with Yiddish speakers in Eastern Europe we develop initial Yiddish language resources that are used for adaptations of speech and language technologies. Our project aims at the development of resources and technologies that can make the entire AHEYM corpus and other Yiddish resources more accessible to not only the community of Yiddish speakers or linguists with language expertise, but also historians and experts from other disciplines or the general public. In this paper we describe the rationale behind our approach, the procedures and methods, and challenges that are not specific to the AHEYM corpus, but apply to all documentary language data that is collected in the field. To the best of our knowledge, this is the first attempt to create a speech corpus and speech technologies for Yiddish. This is also the first attempt to work out speech and language technologies to transcribe and translate a large collection of Yiddish spoken language resources.

Keywords: Yiddish, Speech and Language Corpus, Speech Technology

1. Introduction

Yiddish, in spite of the long tradition of scholarship, can be counted nowadays among languages with an uncertain future. This is – among others – due to its dispersed status and low number of speakers. Ethnologue[1] reported no more than 1.5 mil. speakers of Eastern Yiddish worldwide in 1991 (with some 600,000 speakers in 1991 and 200,000 speakers in the USA estimated by the Modern Language Association, the latter number undated) and 50,000 speakers world-wide for Western Yiddish in year 2000. From the time when this data was collected, the numbers of speakers in Eastern Europe diminished dramatically. What is disappearing is the cultural heritage but also linguistic diversity of the multitude of Yiddish dialects. Another reason for classifying Yiddish as a potentially endangered, and surely under-resourced, is that it belongs to the group of languages with only limited speech and language technologies resources. Crucially, Yiddish recordings stored in archives usually lack word-by-word or sentence-by-sentence transcription, which limits the potential audience that can use and work with the recordings to researchers who can understand Yiddish and – importantly – its dialectal variants. Further, any large scale search and analysis of such collected data is impractical without transcription. In this respect, Yiddish is like hundreds of other endangered languages where researchers in the past decades frantically collected audio recordings of the language but only limited annotation for the recordings is usually available.

In this paper we describe efforts to create an initial set of language resources and tools that later serve as the basis from which we bootstrap annotations and transcriptions for already existing hundreds of hours of Yiddish audio and video recordings. We focus in this project on the resources provided by the Archives of Historical and Ethnographic Yiddish Memories (AHEYM)[2] project.

2. About AHEYM

The original rationale for the project was the language data from the AHEYM collection. The AHEYM is a linguistic and oral history project that includes Yiddish language interviews, with most of speakers born between the 1900s and the 1930s (Kerler, 2014). The corpus encompasses nearly 800 hours of digital video recordings with the speech of more than 400 speakers at over 110 locations. The interviews were conducted primarily in the countries of Eastern Europe: Ukraine, Romania, Moldova, Hungary, Lithuania, Latvia, Slovakia, and Israel, in the years 2002 - 2015, the research being primarily financed by two NEH grants awarded in 2005 and 2009.

The featured material is fascinating linguistically, comprising of texts from a range of closely related dialects, often with Slavic influences. The interviews provide a breadth of information, relating the life stories of Yiddish speakers in Eastern Europe during and after WW2, as well as elements of folklore or local oral history. The interviews include linguistic and dialectological data, oral histories of Jewish life in Eastern Europe, Holocaust testimonials, musical performances (including Yiddish folk songs, liturgical and Hasidic melodies, and macaronic songs), anecdotes, folk narratives, children's ditties, folk remedies, fragments of Purim plays, reflections on contemporary Jewish life in the region, and guided tours by local residents of sites of Jewish memory in the region.

While the interviews address topics of interest for historians, sociologists, ethnographers, they also deliver a breadth of linguistic information. They document and trace dialectological data, which makes it possible to map out the historical development and the geographical distribution of Yiddish dialects, the development of the Yiddish language

[1] See for more details `http://www.ethnologue.com/country/UA/languages`.

[2] See for more details `http://www.iub.edu/~aheym/`.

in general, and the dynamics of interregional connections. From the point of view of theoretical linguistics the interviews provide the basis for the analysis of language change mechanisms in general, and the insights in language typology, language contact, multilingualism, and more. From the perspective of NLP, the data offers a number of interesting challenges, including, among others, code switching, and potentially an adaptation of NLP tools developed originally for one dialect to extend the use for another dialect, or to use speech recognition modules developed for Slavic or Germanic languages in the context of Yiddish.

The entire AHEYM collection is placed with the Indiana University Archives of Traditional Music and its EVIADA (Ethnographic Video for Instruction and Analysis Digital Archive), and it was meant to be transcribed and made available to the public, but after transcribing small parts of the material it became apparent that manual transcription of the complete material is not feasible. AHEYM, like many other documentation projects, has reached the Transcription Bottleneck – only a small proportion of the collected data could be transcribed using the traditional methods, that is, manually. This is so because of the extreme costs of transcription process in terms of time, effort and expertise, where transcription takes 50-100 times the duration of the recording itself, depending on the quality and type of the recordings and the depth of the transcription/annotation. Transcription requires also an expert fluent in the language, technology, and annotation standards at the same time.

The solution to the Transcription Bottleneck problem we pursue is to use corpus and speech and language technology techniques to facilitate and automatize transcription. However, when working on the AHEYM data a number of difficulties became apparent. The recordings in the AHEYM collection are extremely challenging for corpus linguistics and speech and language technologies for various reasons. On the one hand, the recordings were made in various locations and environments that were noisy or that involved multiple speakers interacting and overlapping. The recordings contain code switching between the local language, Russian, Ukrainian, Romanian and Yiddish, or even between German and Yiddish. Some speakers are older and have articulatory difficulties. The quality of the audio in some video recordings varies potentially due to changes in the microphone directions and speaker movements in a multi-party discourse situation. Developing speech and language corpora or technologies from such sources is rather difficult and constitutes a serious research and engineering challenge. In this paper we document the procedures and the solutions to the problems we encountered, and describe the outcomes of the project.

3. Project goals

The Transcription Bottleneck is a common problem for the documentation projects in under-resourced languages. In particular, the archived audio and video materials are associated only with meta information, including the authors of the recording, usually the data about the speakers and the circumstances of the recordings, sometimes – a summary of the content. The recordings usually lack the word-by-word, sentence-by-sentence transcriptions. Without transcription the audio recordings are virtually inaccessible to anybody without the expert knowledge of the recorded language, in our case, without the expert knowledge of Eastern Europe Yiddish dialects. They are not available for further analysis and processing by the majority of researchers or the general public, while the number of people who understand and can process the Yiddish dialects in audio and video recordings is shrinking continuously.

At the same time, resources that are only available as audio or video recordings are not accessible even to the few language experts as they lack fast search or advanced access methods, i.e. research and work with this material is time-consuming itself. Without transcriptions only limited analysis is possible of a small proportion of the already collected data, no search over the total amount of resources or large data analysis is possible. The main goal of the project is to develop speech and language resources and technologies to make the recordings of spoken Yiddish already collected, among others, in the AHEYM collection, more accessible for researchers and general public.

A more general goal is to investigate techniques and approaches for rapid development of speech and language resources and technologies for low-resourced or endangered languages. In particular, we create a reusable infrastructure for language resource development that maximizes the amount of data and resources, at the highest quality of the generated annotations, while minimizing the costs in terms of time and effort invested in the transcription and corpus annotation. The Yiddish speech corpus, thus, feeds a number of other speech corpus projects for low-resourced languages and the research on other languages boosts the development of the Yiddish speech and language resources.

4. Procedure

4.1. Generating the corpus

The initial steps that we took to develop a speech corpus of Yiddish were to select the technically best realized recordings and speakers with clearest pronunciation in the interviews from the AHEYM corpus. The corpus consists of hundreds of hours of digital audio and video recordings. Many interviews contain frequent turn-taking with two or more speakers, or background noise, and other issues that are problematic for speech technology and less useful in speech corpora. We originally selected approximately ten hours of speech for manual transcription and annotation.

During the summer of 2015 we transcribed approximately five hours of spoken Yiddish into four tiers using ELAN (Wittenburg et al., 2006), (Sloetjes and Wittenburg, 2008). The initial tier is Yiddish transcription using Hebrew script. A second tier provided the same transcription using YIVO romanization,[3]then, a phonetic (IPA), and an English translation tier.

Since the transcription of the interviews turned out to be extremely time-consuming and difficult even for initial five hours, we decided to adopt a strategy that we used for developing a spoken corpus for Chatino (Cavar et al., 2016a),

[3]We are using the undecorated *Yidisher visnshaftlekher institut* (YIVO) romanization standard.

this volume, in particular to integrate also the already existing resources – audio files with existing texts that can serve as transcriptions. For the Yiddish corpus, we used texts and audio books from the YIVO Institute for Jewish Research[4] and the Yiddish Book Center[5]. Some texts have been published on these sites as audio books and were read by Yiddish speakers. The advantage of integrating this material into a speech corpus is that it comes with the ready transcription in form of a scanned book. The task that needs to be performed on this kind of resources is to time-align the text with the speech signal, to be able to extract necessary models for Forced Aligners and ASRs.

The integration of the resources from the Yiddish Book Center and the Yiddish Book Center was not always straightforward. For most books we experimented utilizing Optical Character Recognition (OCR) software for Yiddish to be able to convert the scanned books in the archives to text, and to use the resulting text for the transcription and alignment process. Unfortunately, the OCR software has serious difficulties differentiating characters with combining points that are typical for Yiddish.

The initial five hours of recordings from the AHEYM interviews and Yiddish Book Center audio books have been fully time-aligned manually, yielding two tiers in the ELAN transcriptions, Hebrew script Yiddish and romanized YIVO. The first rough time-alignment has been performed on the utterance level using ELAN (Wittenburg et al., 2006). ELAN has proven to be efficient enough to allow for very fast utterance level time alignment and copy-and-paste transcription using existing text, as in our process. In the next step the data is fine-time-aligned at the word level using Praat (Boersma, 2001), (Boersma and Weenink, 2016).

The process of generating corpus with transcriptions and time-alignment using existing text consumes approximately three times real-time, and is very efficient, compared to other methods of speech corpus generation. For comparison, generating initial speech corpora from recordings of freely elicited speech from native speakers consumes significantly more time for the transcription at the raw textual transcription level. Orthographic and phonetic transcription can consume 50 to 100 times real time for some languages and fieldwork data.

To extract the initial IPA for the phonetic transcription tier we used a Yiddish dictionary. Raphael Finkel of the University of Kentucky provides a free web service "The Yiddish Typewriter" (*Di Yidishe Shraybmashinke*) that returns as a web-service from Hebrew script Yiddish – IPA or the romanized YIVO transcript. This service allowed us to generate the transcriptions as well as the phonetic transcriptions in IPA. We also got full access to prof. Finkel's code and digital dictionary for the corpus development purposes, allowing us to perform full-form lookups with scripts.

4.2. Strategies for the forced alignment

In order to train a Forced Aligner from the initial transcribed and time-aligned corpus, we extracted the pairs of audio chunks and corresponding transcriptions at the utterance level. We developed the software tool ELAN2Split[6] to cut utterance level time intervals from specific tiers in ELAN annotation files and to generate a training corpus – audio segments from the WAVE-file and the corresponding transcriptions – as needed by the HTK-based ProsodyLab Aligner[7] (Gorman et al., 2011). Training the Forced Aligner using ProsodyLab requires only ca. 2 hours of spoken language recordings with transcription, and a set of basic language models like a pronunciation dictionary and a description of the phone inventory of the language. The trained Forced Aligner is used to time-align the remaining transcribed material and rapidly expand the volume of the corpus.

In addition to the HTK-based Forced Aligner we also created a Yiddish *text to speech* language model for Espeak (`http://espeak.sourceforge.net/`). Espeak is a *text to speech* engine that is used in Praat to generate audio representations from text that are then aligned in a recording, generating hypotheses about the time intervals for the corresponding transcription text. That is, Praat uses *text to speech* to forced-align transcriptions to an audio recording. We developed a new Yiddish *text to speech* model for Espeak to use it with Praat and the integrated Forced Alignment algorithm (Boersma and Weenink, 2016). Such a model seems to exist already – Yiddish being listed as one of the supported languages in Google Translate (that utilizes also Espeak) – but we could not get access to it.

Using these two different approaches for Forced Alignment allows us to compare the methods, the time investment and output quality, but also to compare the alignment hypotheses and detect deviations or average over both alignments.

The time aligner generates automatic alignments of audio or video recordings and raw textual transcriptions. The output of the Prosodylab Aligner is a Praat-compatible TextGrid file. This automatic time-alignment can be manually validated and corrected using Praat. Since ELAN can import Praat-based TextGrid files, we can perform a full cycle of retraining of the HTK-based Forced Aligner without major effort by regenerating the training corpus with the improved and corrected alignment over a larger corpus.

Given an acoustic and language models we are able to improve the performance of the Forced Aligner and generate an initial Automatic Speech Recognition system.

4.3. Tools and technologies

The AHEYM speech corpus project is part of the GORILLA infrastructure. Within GORILLA we experiment and develop using various speech and language technologies.

For annotation and transcription we use ELAN (Wittenburg et al., 2006). ELAN is an ideal tool for initial utterance level multi-tier annotation and time-alignment. More precise time-alignment at the word level we perform with Praat (Boersma, 2001; Boersma and Weenink, 2016). Praat provides a better visualization of the speech signal based on

[4] See for more details `https://www.yivo.org/`.

[5] See for more details `http://www.yiddishbookcenter.org/`.

[6] See `https://bitbucket.org/dcavar/elan2split` for more details.

[7] See for more details `https://github.com/prosodylab/Prosodylab-Aligner`.

a spectrogram with fine-grained zoom functions that allows for detailed utterance and word-level time-alignment at segment boundaries.

Our own software ELAN2split (https://bitbucket.org/dcavar/elan2split) extracts tuples of time-aligned transcription and audio sequences from ELAN Annotation Files and corresponding WAVE-media files. This ELAN2split software generates a training corpus for e.g. the Prosodylab Aligner from existing ELAN transcriptions. This corpus consists of audio and text-transcription pairs for individual time intervals in the ELAN transcription and time alignment.

As an alternative approach to forced alignment we use Espeak[8] and language specific *text to speech* models. Praat provides a forced alignment functionality that is based on Espeak.

As mentioned above, the difference between Praat-based speech corpus annotation and ELAN-based time-alignment is that Praat allows for much more detailed alignment based on speech signal spectogram views. ELAN does not allow for such a fine-grained annotation. In addition to that, Praat provides a Forced Alignment functionality that is based on Espeak, an open-source text-to-speech engine. Given a speech signal and a corresponding transcription text, Praat generates a speech signal using the text-to-speech engine and, then, hypotheses about the alignment of the generated speech signal and the recorded one, i.e. it utilizes a method that could be described as analysis by synthesis. This method requires the development of models that are mappings from orthographic to phonemic or sound representations that are specific to Espeak.

Espeak and *text-to-speech* models as utilized in Praat not only enable the alignment functionality of Praat for a specific language, but they are also a very interesting documentation tool. The models contain mappings of orthographic representations to the Espeak internal pronunciation symbols and additional exception lexicons. The notation allows for the markup of stress or tone and regularities of stress distribution. A documentation of the acoustic properties of lexical items using this system extends the classical phonetic transcription of lexical lemmata with potential exception lists.

The Prosodylab Aligner is a Hidden-Markov-Model-Toolkit-(HTK)-based Forced Aligner that requires a pronunciation dictionary and the audio transcription pairs on the utterance level to train a model.[9] It generates TextGrid files that are the underlying Praat annotation and time-alignment information file format. These TextGrid files can be opened and edited in Praat (Cavar et al., 2016a).

5. Interoperability and Open Access Philosophy

Language resources are essential for the development of further specific or extended resources and new speech and language technologies. Currently available language resource archives often utilize licenses that impose restrictions on data use. To make the resources and tools available to the potentially broadest audience, first, we have decided to create initial speech corpora using exclusively free resources. Second, we do not impose restrictions on the use of the tools and components we create. The texts that are used in the recordings are distributed freely by e.g. the Yiddish Book Center, and we received a permission to use the recordings and the texts for the purpose of our project. The recordings that we used from the AHEYM corpus are partially published on YouTube and are prepared for public dissemination on the Indiana University website of the Archive of Traditional Music (ATM). The speech corpora and all language models, including parallel corpora and part-of-speech tagged text are made available on the GORILLA infrastructure (http://gorilla.linguistlist.org) under the Creative Commons Attribution-ShareAlike license (Creative Commons, 2016).

Additionally, the project adheres to and promotes best-practice standards in the development of language data and tools. Data are properly annotated in standardized frameworks (e.g. XML-based using annotation standards or common open formats like the ELAN Annotation File format, Praat's TextGrid, RDF, TEI XML), which makes it easier to share and adopt the resources for new applications between different users and projects, and to re-implement technical solutions for different languages.

6. Outcomes and Impact

While one outcome of the project is the first Yiddish speech corpus with additional linguistic annotation, the project has produced a Forced Aligner and acoustic and language models for text-to-speech and automatic speech recognition.

The project gives us insights into ways for optimization of strategies and procedures for speech and language corpora and technology generation, given already existing language resources, e.g. texts, dictionaries and script conversion tools. We experiment with different amounts of data that are necessary to bootstrap a Forced Aligner or an ASR, to be able to determine the optimal point of investment and highest output quality gain that serves our goals, i.e. among others the quick bootstrapping of resources and basic technologies that, in turn, can further facilitate resource development and technology engineering.

Another point is that Yiddish resources we work on have a number of specific linguistic characteristics that are relevant for the development and evaluation of speech technologies. In particular, Yiddish in the recording has a huge dialect-related variation between the individual files. Another challenge is the frequent code-switching between Yiddish and other languages (e.g. German, Russian, Hungarian, Romanian, Hebrew) within individual speakers within the same recording. It is not clear yet from our experiments and settings, how these properties of the recordings might impact the output of the existing speech recognition algorithms.

The speech tools and other language technologies that we provide will potentially make accessible the vast amount of recorded interviews in the AHEYM corpus to the broader

[8] See (http://espeak.sourceforge.net for more details.

[9] See for the Hidden Markov Model Toolkit (HTK) http://htk.eng.cam.ac.uk and Young et al. (2006).

public. We hope that the tools will be useful to also help with the transcription of the audio collection in the Yiddish Book Center and the YIVO Institute for Jewish Research. Most of the audiobooks are available as scanned books as well. Those scans are not textual, but image representations of the scanned book pages. Our technologies and resources can potentially be used to make these resources accessible and searchable. Thus, the project will lead to valorization of archives' existing holdings and rectifying the problem of these resources being under-used.

Since the speech corpus that we have created is translated at least to English, in fact, it constitutes a parallel corpus and can also be used to train machine translation models or help improve the existing translation model for Yiddish in Google Translate. This way we make the invaluable resources in the AHEYM corpus and the Yiddish Book Center available in many more languages and open them up to the audiences which could never use these resources before.

Last but not least, once the interviews are word-by-word transcribed, using the Yiddish transcriptions and/or English translations the big-scale corpus-based analysis is possible – whether for historical, sociological, or linguistic studies – something that is not possible with just raw acoustic data. A transcribed and time-aligned corpus is a treasure full of invaluable data for language-related research. The acoustic and language models contain information about distributional properties of the sounds of the language, which is of great interest to speech scientists. Moreover, once the transcripts are generated, it will become possible to add higher-level linguistic annotations, e.g. for syntax, morphology, pragmatics, discourse analysis, further generate applications useful for language education, etc. The transcripts can be used to search the audio recordings for socio-historical information about the lives, cultures, or historical events of the communities captured therein. Unlocking and mining the cultural wisdom from the past embedded in the linguistic documentation would produce valuable assets to fields like history, sociology, anthropology, psychology, and potentially more.

7. Collaborators

The speech corpus of Yiddish, all models and the documentation of the annotation schema are hosted at GORILLA, an archive and language resource platform at The LINGUIST List and The Archive of Traditional Music, at Indiana University: `http://gorilla.linguistlist.org/` (Cavar et al., 2016b), this volume. GORILLA hosts and develops at the moment speech and language corpora for other low-resourced languages, among them, Eastern Chatino (Cavar et al., 2016a), this volume, Burmese, Baharlu Turkish, and Croatian, which allows for the comparison of approaches and outcomes across projects and further optimization of tools and algorithm.

The exposure of the results of the Yiddish project in this context maximizes the visibility and use of the presented Yiddish speech corpus beyond the community with an explicit interest in resources for Yiddish.

8. Conclusion

One of the major goals of this project was to explore the possibilities to use speech and language technologies to enable access to recorded spoken language in audio and video files. This entails addressing the so called Transcription Bottleneck problem by finding an efficient way to utilize common technologies like Forced Alignment to minimize the time and effort of fully manual transcription and annotation of large collections of language data.

By creating speech and language corpora for low-resourced or endangered languages we enable technologies for these languages that will be able to open up the resources that could not be used by broader public and for bigger research endeavors.

With our dissemination and licensing policies we promote also socially-responsible approaches to data collection and sharing that benefit the research and the speaker communities maximally.

9. Acknowledgments

This project emerged as a spin-off from the NSF-funded AARDVARC project (NSF number 1244713). We are grateful to a large group of colleagues who participated in workshops and conferences organized by the AARDVARC team. We are most grateful for ideas, suggestions, help, and comments from Helen Aristar-Dry, Anthony Aristar, and Douglas H. Whalen, the main team members and organizers of AARDVARC. This particular project on Yiddish – however – was not funded by any funding agency or with third party funds.

The initial transcriptions used for the development of the aligner and further tools were produced within the AHEYM project. The major funding for the AHEYM collection was provided by the National Endowment for the Humanities; Robert A. and Sandra S. Borns Jewish Studies Program at Indiana University; Louis and Sybil Mervis, Danville IL; Dr. Alice Field Cohn and Theodor Cohn, New York. Additional funding and support was provided by the Atran Foundation, Indiana University Arts and Humanities Initiative, Indiana University College Arts and Humanities Institute, Indiana University Multidisciplinary Ventures Fund, Indiana University President Council on International Programs; Indiana University Russian and East European Institute, Institute for Digital Arts and Humanities at Indiana University, The Archives of Traditional Music at Indiana University, The EVIA Digital Archive Project and Vilnius Yiddish Institute.

The creation of the corpus was only possible through the cooperation with the entire AHEYM team at Indiana University. We are in particular grateful to Avi Lang for his expertise and help transcribing and annotating a large part of the corpus. Lwin Moe and Levi King helped us scripting automatic phonetic transcription using resources made available by Raphael Finkel. Raphael Finkel at the University of Kentucky provided us with his Yiddish dictionary that contains lemma, full-forms and IPA transcriptions. The Yiddish Book Center allowed us to use some of their audiobooks for transcription and integration into the corpus. The YIVO Institute for Jewish Research made audio recordings

and transcriptions available online that reduced our work effort to just time-alignment.
We are grateful to the interns and visiting students at LINGUIST List during the summer of 2015 for their help with the corpus creation and annotation process using ELAN, the Prosodylab Aligner, and Praat and Espeak based *text to speech* model creation and experimenting. Many graduate students from the Department of Linguistics at Indiana University have participated and contributed to the resources. In particular, we are grateful to Aaron Albin, Zac Branson, Levi King, and Alec Wolyniec for their help and support in the early phases. All editors and graduate assistants of LINGUIST List – Sara Couture, Andrew Lamont, Lwin Moe, Anna White, Ashley Parker, Amanda Foster – were helpful with their suggestions, comments and other contributions.

10. Bibliographical References

Boersma, P. and Weenink, D. (2016). Praat: doing phonetics by computer [computer program]. Version 6.0.14, retrieved 11 February 2016.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Cavar, D., Cavar, M., and Cruz, H. (2016a). Endangered language documentation: Bootstrapping a Chatino speech corpus, forced aligner, ASR. In *Proceedings of LREC 2016*. ELRA.

Cavar, D., Cavar, M., and Moe, L. (2016b). Global Open Resources and Information for Language and Linguistic Analysis (GORILLA). In *Proceedings of LREC 2016*. ELRA.

Creative Commons. (2016). Creative commons attribution-sharealike 4.0 international. https://creativecommons.org/licenses/by-sa/4.0/, Mar.

Gorman, K., Howell, J., and Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, 39(3):192–193.

Kerler, D.-B. (2014). Surviving remnants of yiddish folksinging and creativity in contemporary ukraine [in yiddish]. *Afn Shvel*.

Sloetjes, H. and Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. In Nicoletta Calzolari, et al., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). Elan: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. C. (2006). *The HTK Book, version 3.4*. Cambridge University Engineering Department, Cambridge, UK.