*Proceedings of ParlaCLARIN II Workshop*, pages 40–43
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

# Challenges of Applying Automatic Speech Recognition for Transcribing EU Parliament Committee Meetings: A Pilot Study

**Hugo de Vos and Suzan Verberne**
Institute of Public Administration and Leiden Institute of Advanced Computer Science, Leiden University
h.p.de.vos@fgga.leidenuniv.nl, s.verberne@liacs.leidenuniv.nl

### Abstract

We tested the feasibility of automatically transcribing committee meetings of the European Union parliament with the use of Automatic Speech Recognition techniques. These committee meetings contain more valuable information for political science scholars than the plenary meetings since these meetings showcase actual debates opposed to the more formal plenary meetings. However, since there are no transcriptions of those meetings, they are a lot less accessible for research than the plenary meetings, of which multiple corpora exist. We explored a freely available ASR application and analysed the output in order to identify the weaknesses of an out-of-the box system. We followed up on those weaknesses by proposing directions for optimizing the ASR for our goals. We found that, despite showcasing acceptable results in terms of Word Error Rate, the model did not yet suffice for the purpose of generating a data set for use in Political Science. The application was unable to successfully recognize domain specific terms and names. To overcome this issue, future research will be directed at using domain specific language models in combination with off-the-shelf acoustic models.

**Keywords:** Automatic Speech Recognition, European Union Parliament, Political Science

## 1. Introduction

The plenary meetings of the European Parliament have been a thankful study object (Glavaš et al., 2019; Hollink et al., 2018) both in Natural Language Processing (NLP) and in Political Sciences (Greene and Cross, 2017). The recordings and transcripts of these meetings contain rich information about the decision processes of the EU. This research is facilitated by the availability of the data: A large quantity of speeches is available in a standardized format that is relatively easy to come by. For example via the LinkedEP project (Aggelen van et al., 2017) or the Europarl corpus (Koehn, 2005).

However, for political scientists, the relevance of these data is limited, because of the way the plenary meetings are structured. Plenary meetings consist of short (often one-minute) speeches that are mostly read from paper. Such speeches are well prepared and thought out, and speaking times are very limited. There is no room for any interruptions or other means to directly react to what is happening on the floor. In other words: there is no actual debate. The plenary sessions of the EU can be considered a case of formal language use, as opposed to spontaneous speech. Because of their limited length and the formal language use, the speeches only contain superficial information about the topics discussed and very limited information about the position of the meeting participants towards these topics.

Much more interesting sources of information about processes in the EU parliament are the EU parliament *committees*, in which matters are discussed on a more technical level. These domain specific meetings are where the actual debate takes place: specific issues are debated with more detail than in the plenary meetings. EU committees consist of Members of the European Parliament (MEPs) and are centered around core topics within the EU, such as Civil Liberties, Justice and Home Affairs (LIBE committee) or Internal Market and Consumer Protection (IMCO).[1]

The problem of the meetings of the committees is that, contrary to the plenary meetings, there are no transcriptions of those meetings. Only a coarse agenda and minutes are available. This severely limits how well this data can be used for political research, because it is only possible to find the meetings based on metadata. Moreover, listening to the audio files is time-consuming and therefore unfit for any larger scale research.

It is for this reason that we set up this project of transcribing the committee meetings using Automatic Speech Recognition (ASR) with the aim of creating a corpus that is relevant to political researchers. This will make large scale analyses of this rich data collection possible, and allows for political research that goes beyond the available but shallow plenary meetings.

In this paper we discuss results obtained in a pilot on the automatic transcription of the EU parliamentary committee meetings. We studied whether it is feasible to automatically transcribe committee meetings with sufficient quality to be used in political research.

## 2. Data

The input data consist of audio recordings of the meetings of EU parliamentary committees that were retrieved from the official database of recorded meetings.[2]
The recordings are (almost) entirely in English. If a speaker talks English, their audio recording is directly recorded in the audio file. If a speaker does speak an other language than English, then the first few words of the speaker can be

[1] For a full list of EU parliamentary committees consult https://www.europarl.europa.eu/committees/en/parliamentary-committees.html

[2] https://www.europarl.europa.eu/ep-live/en/committees/search

heard after which the sound is overlayed with the recording of an interpreter speaking in English.

### 2.1. Data sampling

We selected the audio recordings of 5 random meetings of the LIBE committee of the EU parliament. In total these 5 recordings had a length of 9 hours and 51 minutes. Of those 5 meetings we manually transcribed the first 15-20 minutes as reference material, leading to 1 hour and 20 minutes of manually transcribed audio which constituted a total of 7902 words. An average of 100 words per minute might seem low, but this is due to a lot of silences, for example when a new speaker needs to walk to the microphone.

### 2.2. Manual transcription

Manual transcription for the purpose of ASR evaluation was not done from scratch but by editing the output of the ASR-application (See Section 3.1. for description of the application). We realize that this way of transcribing might insert biases in the transcription. Yet for the purpose of this pilot we think it does suffice. During the transcription a few special elements were used:

- If a section of the audio was inaudible or the exact thing said could not be understood this was transcribed as ****.

- The parts where the first few words of a non English speaking speaker were heard before the sound of the entire sequence before interpreter started was described as a single instance of **FOREIGN**, since we were no to recognize the different words in a foreign language.

## 3. Methods

### 3.1. ASR application

For transcribing the sound files we used the English Automatic Speech Recognition Webservice (based on the Kaldi framework), developed at the University of Twente, version 0.1[3]. This webservice is based on the KALDI framework for speech recognition (Povey et al., 2011). It was trained on the TED-LIUM data set[4], which is a data set of transcribed TED-talks. This is especially beneficial for our case since the model is able to dealing with non native speakers of English.

### 3.2. Evaluation

To evaluate the quality of the ASR for our data we used the Word Error Rate (WER) on our manually transcribed sample. This is a common metric for measuring and comparing the quality of ASR-systems (Chiu et al., 2018; Toshniwal et al., 2018). The WER is based on the Levenshtein distance, which defines the minimal edit distance between two strings. For this minimal edit distance, the number of insertions, deletions and substitutions divided by the total number of words:

$$\mathrm{WER} = \frac{\mathrm{Substitutions} + \mathrm{Deletions} + \mathrm{Insertions}}{\mathrm{Total\ nr\ of\ words}} \qquad (1)$$

## 4. Results

The results are summarized in Table 1. The table indicates that the ASR quality is reasonably good with WERs between 5 and 14. We think that these results provide us with a decent baseline, especially given the specificity of the domain (see Section 5.) and the fact that all of the speakers in our sample were non-native speakers of English and that the model was not tuned for this. As a comparison, the WER reported for end-to-end ASR (Hadian et al., 2018) on the 300-hour Switchboard corpus is 9.3; the best WER reported for the LibriSpeech test-other set is 5.0 and for the LibriSpeech test-clean set it is 2.2.[5]

Table 1 also shows that the WER differs quite largely between the five topics, with the Justice scoreboard being the most difficult to transcribe. An inspection of the output indicates that this is mainly due to variation in speaker pronunciation (some speakers are easier to understand than others), and the amount of technical language, which is particularly low in the topic ‘Appointment of vice-chairs’.

| **Topic** | **# transcribed words** | **WER** |
|---|---|---:|
| Appointment of vice-chairs | 2294 | 5.15 |
| Justice scoreboard | 1092 | 13.19 |
| Visa code | 1815 | 6.61 |
| Eurojust evaluation | 1442 | 9.43 |
| Legal Assitance for MEPs | 1259 | 8.11 |

Table 1: WER values for the five manually transcribed texts

## 5. Error Analysis

Although the WERs presented above are a promising start, there are some issues that could seriously limit the usefulness of the transcriptions for political research and need attention by future ASR developments for this domain.

### 5.1. Person Names

In order for the data to be relevant for political research, the correct recognition of person names is essential. A lot of political science research is centered around questions involving people, roles, parties, and their contributions (Simaki et al., 2018). However, the recognition of names is a challenge for our generic ASR system.

In total 48 times a name was mentioned in our sample, yet only one time the name was recognized correctly, which was in the following context:

*Dear colleagues dear friends. Dear <u>Alexander.</u>*

[3] The webservice is available from https://webservices-lst.science.ru.nl

[4] https://www.openslr.org/19/

[5] Results on these tasks are listed on https://paperswithcode.com/task/speech-recognition

Examples of incorrectly recognized names were more abundant. For example in:

> *This is one of the greatest successes of this committee of its secretariat and Emilio De Capitani sitting next to me the head of it.*

which was recognized as:

> *This is one of the greatest successes of this committee of its secretariat in dick up a county sitting next to me the head of it.*

What becomes clear from this example is that apart from the name every word was recognized correctly. However, without a correctly recognized name, this sentence becomes close to incomprehensible, let alone useful for political research.

Another example where a name was miss-recognized is:

> *We have for the rapporteur Axel Voss among us.*

which was recognized as:

> *We have what the rapporteur access among us.*

In this example also words besides the name Axel Voss are misrecognized, yet only changing the word *access* for *Axel Voss* will make the sentence more or less understandable, showing again the detrimentality of correctly recognizing names.

### 5.2. Institution names

Apart from person names, also institution names appear to be hard to recognize. One of the transcribed fragments is about the Eurojust[6] institution. Eurojust was mentioned a total of 12 times in 1442 transcribed words yet it was never recognized correctly. For example

> *(...) and present the Eurojust annual report two thousand and thirteen (...)*

was recognized as

> *(...) and presenter you just annual report two thousand and thirteen (...)*

### 5.3. Domain-specific terms

Another type of words that is often not recognized are specific jargon terms from the EU. One of the transcribed fragments was about new legislation regarding EU visas: the visa code. Although the word *visa* was correctly recognized 14 out of 35 times, the phrase *visa code* was only correctly identified 1 out of 12 times.
This again forms a problem for political science research. If one of the most important terms of a documents is not recognized properly this will affect any further analysis on the content of the meetings.

## 6. Next steps

We presented the results from a pilot studying the feasibility of automatically transcribing EU committee meetings for the sake of political research. We are following-up on this work on two directions: adapting ASR, and downstream analysis of the resulting transcripts.

### 6.1. Improving the ASR

Modern ASR-systems consist of three or four parts: an acoustic model, a language model a decoder and often a vocabulary. The latter is sometimes implicit to the language model. The acoustic model is typically a deep neural network that is trained to map an acoustic signal to symbols (either graphemes or phonemes) representing the sounds. The output of this model is a string of symbols that is transcribed into a sequence of words by the decoder. This decoder compares the string of characters with a language model. A language model contains information about what the probability of occurrence is of words and word sequences (Chan et al., 2016). Based on the information of the acoustic model and the language model, the decoder determines what word is the most likely (Synnaeve et al., 2019).

Adapting an ASR-pipeline towards a specific domain can be done in all the components: the acoustic model, the language model, the vocabulary and the decoder.

For our application we will achieve the domain adaptation by adding a domain-specific language model and vocabulary. The reason for this is that most errors can be ascribed to out-of-vocabulary terms. If a term is not in the language model or underrepresented in the language model, an ASR-system will be unable to recognize it.[7] In the case of names of persons and institutions, most of them will not have been present in the corpus that a generic language model is trained on. The result is that the decoder will not consider those words when analyzing the output of the acoustic model. A word such as *visa* might be in the training corpus of a generic ASR language model, but it will be less common than in the EU-domain and also occur in different contexts within the EU than outside the EU. Therefore it will also be more often disregarded as the most likely term.

We can leverage the availability of written documents from the European Union to train a domain-specific language model. For example, all the transcripts from plenary meetings of the EU parliament can be used for this purpose, since they are readily available: for example via the LinkedEP project (Aggelen van et al., 2017).

### 6.2. Analysis of the transcripts

Once we have a collection of reliable transcripts of the meetings of the EU parliament committees, we plan to ex-

[6] European Union Agency for Criminal Justice Cooperation.

[7] The alternative, training a domain-specific, acoustic model would require large amounts of time and resources: hundreds of hours of transcribed acoustic data would be needed, together with substantial computational power (Synnaeve et al., 2019).

plore a number of interesting research directions. In this section we will paint some of the possibilities.

**Opinion mining and stance analysis** We plan to use the data set to mine the opinions and standpoints of different MEPs over time. This falls within a tradition in Political Science of measuring positions of actors based on texts. This can either be a position towards a specific subject (Lopez et al., 2017) or in the larger political spectrum, for example on a right-left scale (Lowe et al., 2011).
Such analyses can be made extra interesting in the case of this particular data. Members of these committees are not only affiliated to their committees, but also to national political parties, European political fractions and in some extent also to their home country. Linking the textual database to other databases holding these affiliations will add interesting dimensions that can provide for exciting new research.

**Topic modelling** Other research possibilities with this data set would include (dynamic) topic modeling (Blei and Lafferty, 2006). It would be a novel research direction to explore what the main topics are prevalent within and between committees over time.

## 7. Conclusion

In this paper we explored the possibility of generating a corpus of transcriptions of EU parliament committee meetings using a generic ASR system. We conclude that the system we used shows promising results, yet does not suffice. However, we deem it possible to make adaptations towards a working system. The main problems are recognizing names and domain-specific terms that are outside the vocabulary of a generic system. For this reason, our next steps are to train a domain-specific language model and vocabulary, leveraging the large amount of written EU documents available. Our long-term aim is to enable researchers in the field of political science and public administration to better analyze the EU policy processes with the help of automated text analyses.

## 8. Bibliographical References

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120.

Chan, W., Jaitly, N., Le, Q., and Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4960–4964. IEEE.

Chiu, C.-C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R. J., Rao, K., Gonina, E., et al. (2018). State-of-the-art speech recognition with sequence-to-sequence models. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4774–4778. IEEE.

Glavaš, G., Nanni, F., and Ponzetto, S. P. (2019). Computational analysis of political texts: Bridging research efforts across communities. In *Proceedings of the 57th annual meeting of the association for computational linguistics: Tutorial abstracts*, pages 18–23.

Greene, D. and Cross, J. P. (2017). Exploring the political agenda of the european parliament using a dynamic topic modeling approach. *Political Analysis*, 25(1):77–94.

Hadian, H., Sameti, H., Povey, D., and Khudanpur, S. (2018). End-to-end speech recognition using lattice-free mmi. In *Interspeech*, pages 12–16.

Hollink, L., van Aggelen, A., and van Ossenbruggen, J. (2018). Using the web of data to study gender differences in online knowledge sources: the case of the european parliament. In *Proceedings of the 10th ACM Conference on Web Science*, pages 381–385.

Lopez, J. C. A. D., Collignon-Delmar, S., Benoit, K., and Matsuo, A. (2017). Predicting the brexit vote by tracking and classifying public opinion using twitter data. *Statistics, Politics and Policy*, 8(1):85–104.

Lowe, W., Benoit, K., Mikhaylov, S., and Laver, M. (2011). Scaling policy preferences from coded political texts. *Legislative studies quarterly*, 36(1):123–155.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., and Vesely, K. (2011). The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, December. IEEE Catalog No.: CFP11SRW-USB.

Simaki, V., Paradis, C., and Kerren, A. (2018). Evaluating stance-annotated sentences from political blogs regarding the brexit: a quantitative analysis. *ICAME Journal*, 42(1).

Synnaeve, G., Xu, Q., Kahn, J., Grave, E., Likhomanenko, T., Pratap, V., Sriram, A., Liptchinsky, V., and Collobert, R. (2019). End-to-end asr: from supervised to semi-supervised learning with modern architectures. *arXiv preprint arXiv:1911.08460*.

Toshniwal, S., Kannan, A., Chiu, C.-C., Wu, Y., Sainath, T. N., and Livescu, K. (2018). A comparison of techniques for language model integration in encoder-decoder speech recognition. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 369–375. IEEE.

## 9. Language Resource References

Aggelen van, Astrid and Hollink, Laura and Kemman, Max and Kleppe, Martijn and Beunders, Henri. (2017). *The debates of the european parliament as linked open data*. IOS Press.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86. Citeseer.