

Quantitative Analysis of Editing in Transcription Process in Japanese and European Parliaments and its Diachronic Changes

Tatsuya Kawahara

School of Informatics, Kyoto University
Sakyo-ku, Kyoto 606-8501 Japan
kawahara@i.kyoto-u.ac.jp

Abstract

In making official transcripts for meeting records in Parliament, some edits are made from faithful transcripts of utterances for linguistic correction and formality. Classification of these edits is provided in this paper, and quantitative analysis is conducted for Japanese and European Parliamentary meetings by comparing the faithful transcripts of audio recordings against the official meeting records. Different trends are observed between the two Parliaments due to the nature of the language used and the meeting style. Moreover, its diachronic changes in the Japanese transcripts are presented, showing a significant decrease in the edits over the past decades. It was found that a majority of edits in the Japanese Parliament (Diet) simply remove fillers and redundant words, keeping the transcripts as verbatim as possible. This property is useful for the evaluation of the automatic speech transcription system, which was developed by us and has been used in the Japanese Parliament.

Keywords: Parliamentary record, Japanese Diet, European Parliament

1. Background

Transcription is a process of converting speech into text, and there are two goals: accuracy, or faithfulness to speech, and readability, or easiness of reading. They are often in trade-off relationships. Thus, standards or guidelines on transcription, including editing, have been strictly designed and enforced in Parliamentary reports compared with private sectors. One of the well-known is the Hansard of the British Parliament (Mollin 2007). They are, however, different across languages and countries and also change over time. They may be affected by other factors such as TV broadcasting, SNS, and the use of automatic speech recognition (ASR) technology. In this study, a quantitative analysis of the editing process is conducted for Japanese and European Parliamentary meetings by comparing the faithful transcripts of audio recordings and the official transcript records.

2. Edits in Transcription Process

An example of a faithful transcript and an official record is shown in Figure 1. There are many factors requiring edits in the transcription process. First of all, disfluency must be removed. Other kinds of redundancy need to be removed. Then, grammatical errors must be corrected. Some colloquial expressions should be rephrased into formal expressions. Last but not least, speech does not have explicit punctuations, unlike text, so it is necessary to insert periods and commas in appropriate places. In addition to these edits, some structural modifications are sometimes made to improve readability. Moreover, some semantic corrections are made for apparent mistakes. These are explained one by one.

2.1 Removal of Redundancy

Fillers, such as “um” and “ahh” in English, must be definitely removed. They are not transcribed by human stenographers in the first place. They can also be automatically eliminated by ASR systems.

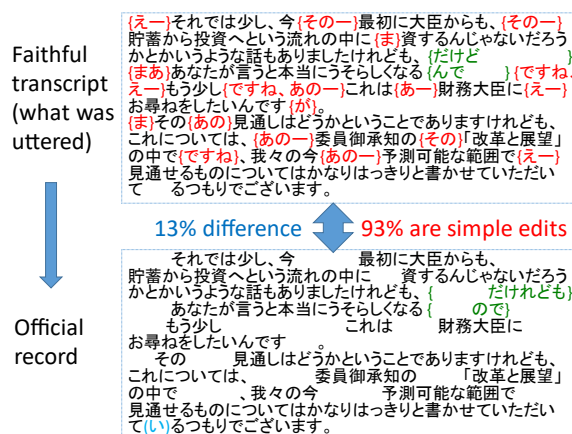


Figure 1: Example of a faithful transcript and an official record in the Japanese Diet

Repeats and repairs must also be removed, but their automatic removal is difficult.

Discourse markers, such as “OK” and “yes” in English, can be kept, but too many tokens reduce readability. Other extraneous expressions, such as “Thank you,” can also be kept, but the removal of them would improve readability.

2.2 Correction of Errors and Colloquial Expressions

There are some kinds of grammatical errors whose correction is mandatory, for example, missing or incorrect articles such as “a” and “an,” and improper use of prepositions such as “in” and “on.” Some kinds of colloquial expressions should also be corrected; for example, “was like” changed to “said” and “but” changed to “however.” But we note language use changes over time. Handling of dialect is also an issue. While some dialects cannot be understood by many readers, dialect is often used to express the identity of the speaker.

2.3 Structural and Semantic Corrections

Some structural reordering is conducted; for example, “Finnish incoming presidency” is changed to “incoming Finnish presidency.” It is often necessary to split a long sentence into a sequence of plain sentences.

On the other hand, semantic correction needs attention. While apparent errors such as mistakes of “billion” and “million” should be corrected via a proper process, it is a question whether errors of proper names or fact errors should be corrected because MPs should be responsible for their statements. Especially when the errors affected the following interaction in the meeting, they should not be corrected.

3. Corpus Analysis in European Parliament and Japanese Diet

3.1 Used Corpora

A corpus-based analysis was conducted using transcripts from the European Parliament (Koehn 2005) and the Japanese Diet (the House of Representatives) (Akita 2006). From the European Parliament proceeding, English-speaking parts in some plenary sessions in 2007 were selected. With regard to the Japanese Diet, a number of sessions in committee meetings held during 2005-2007 were selected. They were selected to cover all major meetings in a good balance. In addition to the official proceeding text, faithful transcripts of spoken words, including fillers and disfluencies, were manually prepared for the analysis. In fact, these faithful transcripts were prepared for the development of the ASR system. The general statistics of the two corpora are shown in Table 1.

The overall edit distance in words between the faithful transcript and the official record is approximately 13% in the Japanese Diet, while it is over 20% in the European Parliament. The larger difference in the European Parliament is attributed to grammatical strictness in the English language compared to Japanese. For example, a subject and prepositions are often omitted in Japanese, while they cannot be omitted in English. There are also many non-native English speakers in the European Parliament.

3.2 Analysis of Edit Categories

Table 2 lists the statistics of edit categories described in the previous section. A large majority (93%) of edits in the Japanese Diet are simple and can be classified as deletion, insertion, or substitution (correction) of words. Almost 90% of them are deletions, and almost half of them involve filler removal. On the other hand, there are much more complex corrections in the European Parliament because English needs many grammatical corrections and syntax reordering. Thus, there is a different tendency according to the language.

Here are typical edit patterns observed for English in the European Parliament. The most frequently removed words other than fillers are “thank you,” “I

think,” and “also,” while the most frequently inserted ancillary and functional words are “the,” “that,” “a,” “also,” and “and.” The most frequently corrected patterns are “but → however,” “thank you → Mr.,” “would → should,” “our → the,” and “this → that.”

Table 1 General statistics of corpora

	European Parliament	Japanese Diet
#words (faithful transcript)	30.9K	418K
#words (official record)	27.1K	379K
% of edited words	20.5%	12.9%

Table 2 Statistics of edit categories

		European Parliament	Japanese Diet
Remove	Fillers	11.6%	46.7%
	Repeats/repairs	11.0%	9.4%
	Discourse markers	1.8%	18.4%
	Extraneous expressions	16.8%	3.0%
Correct	Grammatical errors	20.1%	7.5%
	Colloquial expressions	18.0%	8.4%
Reorder		19.6%	5.9%

3.3 Analysis per Meeting Category and Diachronic Changes

The occurrence ratio of edits per committee in the Japanese Diet is shown in Figure 2. There was a tendency in 2007 that fewer edits were made in the Commission on the Constitution, the Committee on Budget, and the Question Time. While one-on-one interaction is a norm in other committees, the Commission on the Constitution adopts the style of free discussions by all members. This style affects the transcription process. The Committee on Budget and the Question Time are usually broadcasted on the national TV channel, and this may affect the editing process.

In Figure 2, we can also observe a significant change from 2007 to 2016. The ratio of edits has been reduced by 40% over the ten years.

4. Discussions

There are several causes for the reduction of edits. Most significantly, phrase reordering is not done anymore. Some discourse markers are now kept, and some repeats are allowed, such as those expressing emphasis. Moreover, many colloquial expressions are getting accepted. These suggest that the transcripts become more verbatim than before.

There are some possible reasons for this trend. First is the deployment of the ASR system since 2011. Reporters now edit a faithful transcript, which is

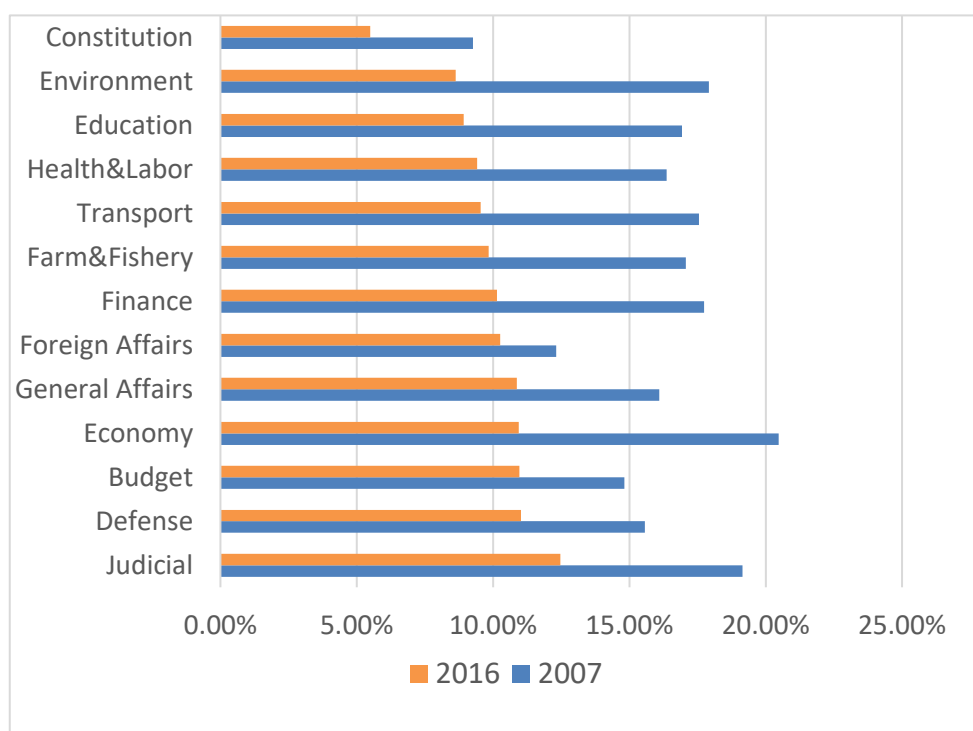


Figure 2: Statistics of edits per meeting category in the Japanese Diet in 2007 and 2016

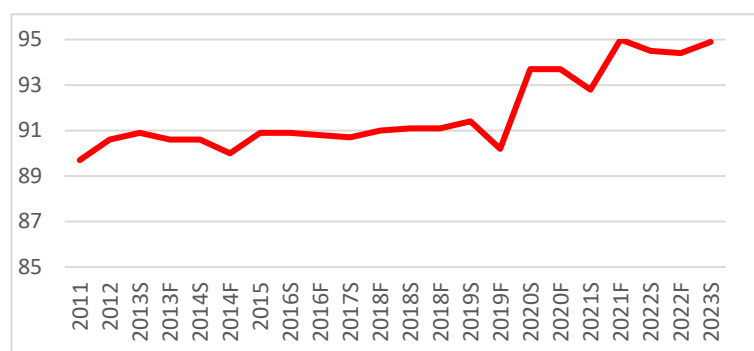


Figure 3: Character correct rate (%) of automatic speech recognition for Japanese Diet

generated by the ASR system. In the old system based on stenography, they typed in text with editing in mind. The second factor is Internet broadcasting. All meetings are broadcast via the Internet. They are also archived and can be accessed at any time; thus, they can be referred to on social media. With these factors, the guidelines by editors may have been changed, although there is no written guideline in the Japanese Diet.

Besides these factors, it is pointed out that there has been a global trend toward writing more verbally or in a closer way to speech in recent decades, even in parliaments (Korhonen 2023).

5. Evaluation of Automatic Transcription System

Since 2011, the House of Representatives of Japan has adopted the ASR system, which was developed by the author's lab (Kawahara 2012). The acoustic model was trained with 1000 hours of parliamentary

speech, and the language model was trained with 25 years of meeting records.

It is found that a large majority of edits for the transcripts of the Japanese Diet are the removal of fillers and discourse markers. This property makes it easy to automatically evaluate the performance of the ASR system without preparing the faithful transcripts. The word/character correct rate is defined by the edit distance minus insertion errors, which counts only substitution and deletion errors. Notice again that a majority of insertions in automatic transcription are due to fillers and redundant words, which must be omitted in the final transcript. The effect of other kinds of edits is smaller than 1%.

The character correct rate for each session/year is plotted in Figure 3. It had been steady at around 91% before the ASR system was improved by adopting a deep learning model in 2020, which significantly improved the accuracy to 95%.

6. Conclusions and Future Perspectives

As the ASR system shows very useful performance, the next step will be to automate the post-editing process. This study has been conducted before with many approaches (Charniak 2001, Honal 2003, Hori 2003, Maskey 2006, Neubig 2012, Shitaoka 2004, Yeh 2006), but it has never met the satisfactory level required by the Parliament.

However, recent large language models such as GPT-4 show the functionality of cleaning transcripts either by human or ASR systems. It is a time to revisit the problem.

7. Acknowledgments

This study is supported by JST A-STEP JPMJTR23RD.

8. Bibliographical References

- Mollin, S. (2007). The Hansard hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora* 2(2): 187--210.
- Akita, Y., and Kawahara, T., (2010). Statistical transformation of language and pronunciation models for spontaneous speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 18 (6), 1539–1549.
- Charniak, E. and Johnson, M. (2001). Edit detection and parsing for transcribed speech. *Proceedings of NAACL*.
- Honal, M. and Schultz, T., (2003). Correction of disfluencies in spontaneous speech using a noisy-channel approach. *Proc. EuroSpeech*, pp. 2781–2784.
- Hori, T., Willett, D., and Minami, Y., (2003). Paraphrasing spontaneous speech using weighted finite-state transducers. In: *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*.
- Kawahara, T. (2012). Transcription system using automatic speech recognition for the Japanese Parliament (Diet). In *Proc. AAAI/IAAI*, pp.2224—2228.
- Koehn, P. (2005): Europarl: A Parallel Corpus for Statistical Machine Translation, *Proc. MT Summit*, pp. 79–86.
- Maskey, S., Zhou, B., and Gao, Y. (2006). A phrase-level machine translation approach for disfluency detection using weighted finite state transducers. *Proceedings of InterSpeech*, pp. 749–752.
- Neubig, G., Akita, Y., Mori, S., and Kawahara T. (2012). A monotonic statistical machine translation approach to speaking style transformation. *Computer Speech and Language*, Vol.26, No.5, pp.349--370, 2012.
- Shitaoka, K., Nanjo, H., and Kawahara, T. (2004). Automatic transformation of lecture transcription into document style using statistical framework. *InterSpeech*, pp. 2169–2172.
- Yeh, J.-F. and Wu, C.-H. (2006). Edit disfluency detection and correction using a cleanup language

model and an alignment model. *IEEE Transactions on Audio, Speech, and Language Processing* 14 (5), 1574–1583.

Korhonen, T., Kotze H., and Tyrkkö J. eds. (2023). *Exploring language and society with big data: Parliamentary discourse across time and space*. John Benjamins.

9. Language Resource References

- Koehn, P. (2005): [European Parliament Proceedings Parallel Corpus 1996-2011](https://www.statmt.org/europarl/) (<https://www.statmt.org/europarl/>)
- Proceedings of Japanese Diet. [国会会議録検索システム](https://kokkai.ndl.go.jp/#/) (<https://kokkai.ndl.go.jp/#/>)