Multilingual Simultaneous Sentence End and Punctuation Prediction

Ricardo Rei
Unbabel
INESC-ID
Instituto Superior Técnico
ricardo.rei@unbabel.com

Fernando Batista
INESC-ID
ISCTE - Instituto Universitário de Lisboa
fernando.batista@inesc-id.pt

Nuno M. Guerreiro
Instituto de Telecomunicações
Instituto Superior Técnico
nuno.s.guerreiro@tecnico.pt

Luisa Coheur
INESC-ID
Instituto Superior Técnico
luisa.coheur@inesc-id.pt

Abstract

This paper describes the model and its corresponding setup, proposed by the Unbabel & INESC-ID team for the 1st Shared Task on Sentence End and Punctuation Prediction in NLG Text (SEPP-NLG 2021). The shared task covers 4 languages (English, German, French and Italian) and includes two subtasks: subtask 1 – detecting the end of a sentence, and subtask 2 – predicting a range of punctuation marks. Our team proposes a single multilingual and multitask model that is able to produce suitable results for all the languages and subtasks involved. The results show that it is possible to achieve state-of-the-art results using one single multilingual model for both tasks and multiple languages. Using a single multilingual model to solve the task for multiple languages is of particular importance, since training a different model for each language is a cumbersome and time-consuming process. Finally, the code for the shared task is publicly available for reproducible purposes at https://github.com/Unbabel/caption/tree/shared-task.

1 Introduction

The text produced by a speech recognition system or by an automatic machine translation system often includes misplaced punctuation and, in the case of a speech recognition system, the output often consists of raw single-case words, without punctuation marks, and may not even include sentence boundaries. Detecting the sentence boundaries and the missing punctuation in such automatically generated texts improves the quality of such texts, and is often relevant for a number of downstream tasks, such as parsing, information extraction, dialog act modeling, Named Entity Recognition (NER), and summarization (Zechner, 2002; Huang and Zweig, 2002; Kim and Woodland, 2003; Ostendorf et al., 2005; Jones et al., 2005; Makhoul et al., 2005; Shriberg, 2005; Matusov et al., 2006; Peitz et al., 2011; Cattoni et al., 2007; Ostendorf et al., 2008; Liao et al., 2020).

Most of the available studies focus on *full stop* and *comma*, which have higher corpus frequencies, and a number of more restricted studies also consider the *question mark*. However, several punctuation marks can be considered for automatically generated texts, including: *comma*; *period* or *full stop*; *exclamation mark*; *question mark*; *colon*; *semicolon;* and *quotation marks*. Nevertheless, most of these marks rarely occur and are quite difficult to insert or evaluate. Quotations and semicolons, for example, are often used inconsequently and in a highly variable way.

This paper proposes a multilingual model that is able to detect sentence boundaries and predict a wide range of punctuation marks, based on pretrained contextual embeddings. Our architecture is composed of three main building blocks: a pretrained Transformer-based encoder model, an attention mechanism over the encoder layers, and the task classification heads. The proposed model derives from the multilingual model proposed by (Guerreiro et al., 2021), which achieves fairly competitive results in a multi-language scenario, even surpassing the existing results for some of the languages.

The reminder of the paper is organized as follows: Section 2 presents an overview of the related work. Section 3 overviews the data used for training fine-tuning our model. Section 4 presents the building blocks of the model architecture, and the setup parameters. Section 5 reports the experiments

performed and Section 6 presents the corresponding results. Finally, Section 7 presents the most relevant conclusions, and mentions possible future directions.

2 Related work

Proper identification of sentence boundaries and punctuation recovery are two profoundly connected tasks that can result in great improvements for speech processing downstream task (Harper et al., 2005; Mrozinsk et al., 2006; Ostendorf et al., 2008). For that reason, recovering structural information from text produced by Automatic Speech Recognition (ASR) becomes an objective of many studies. Early studies used a combination of n-grams with prosodic classifiers through the general Hidden Markov Models framework (Beeferman et al., 1998; Christensen et al., 2001; Kim and Woodland, 2001). With the development of Conditional Random Fields (CRF) and Maximum Entropy models, researchers were able to successfully improve these task (Huang and Zweig, 2002; Liu et al., 2005, 2006; Batista et al., 2007, 2008, 2009; Lu and Ng, 2010; Batista et al., 2010, 2012; Ueffing et al., 2013).

Regarding machine translation, it is a well-known fact that punctuation and capitalization errors are a predominant problem for Statistical Machine Translation (SMT). Several studies tried to enrich the SMT output by inserting proper capitalization and punctuation in the returned translation (Cattoni et al., 2007; Peitz et al., 2011). Even with Neural Machine Translation (NMT), the punctuation errors are still the most predominant type of errors. Indeed, these represent around 20% of the errors produced by the high performing systems from WMT20 News Translation shared task (Freitag et al., 2021).

Most of the recent approaches for punctuation restoration are based on neural networks such as Recurrent Neural Networks (RNN) and Transformers. With that said, most works treat the problem either as a sequence-to-sequence or as a sequence labelling task (Tilk and Alumäe, 2015, 2016; Che et al., 2016; Klejch et al., 2017; Yi and Tao, 2019; Kim, 2019). Following the recent trends in Natural Language Processing (NLP) some of these works take advantage of pre-trained models such as BERT Cai and Wang (2019); Makhija et al. (2019); Guerreiro et al. (2021). Our shared task participation is mostly based on the work by (Guerreiro et al., 2021) that showed that having one single multilingual model is competitive with having one model trained for each language.

3 Corpora

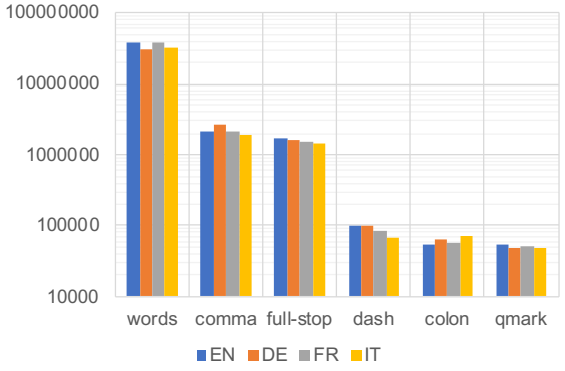

Figure 1: Frequency of each punctuation mark

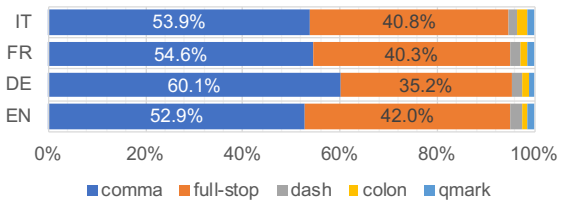

Figure 2: Frequency of each punctuation mark

The SEPP-NLG challenge adopted the Europarl corpus, covering English, German, French, and Italian. The corpus was previously processed in order to remove punctuation marks and case information, as a way to simulate Natural Language Generated text. The challenge considers 5 different punctuation marks: comma (,), full-stop (.), dash (-), colon (:) and question marks (?). Figures 1 and 2 show the frequency of the words and punctuation marks for each one of the languages, considering the training and development sets. As expected, from all the punctuation marks being considered, *comma* is the most frequent, occurring between 52.9% (EN) and 60.1% (DE) of the times, followed by *full-stop*, occurring between 42% (EN) and 35.2% (DE) of the times. All the other punctuation marks into consideration, occur less than 0.24% of the times for all the considered languages. About 95% of the sentences contain between 3 and 50 words, but the maximum sentence length is 303 words for EN, 450 for DE and IT, and 423 for FR. 99% of the sentences contain 1 to 7 punctuation marks, including the corresponding sentence boundary. However,

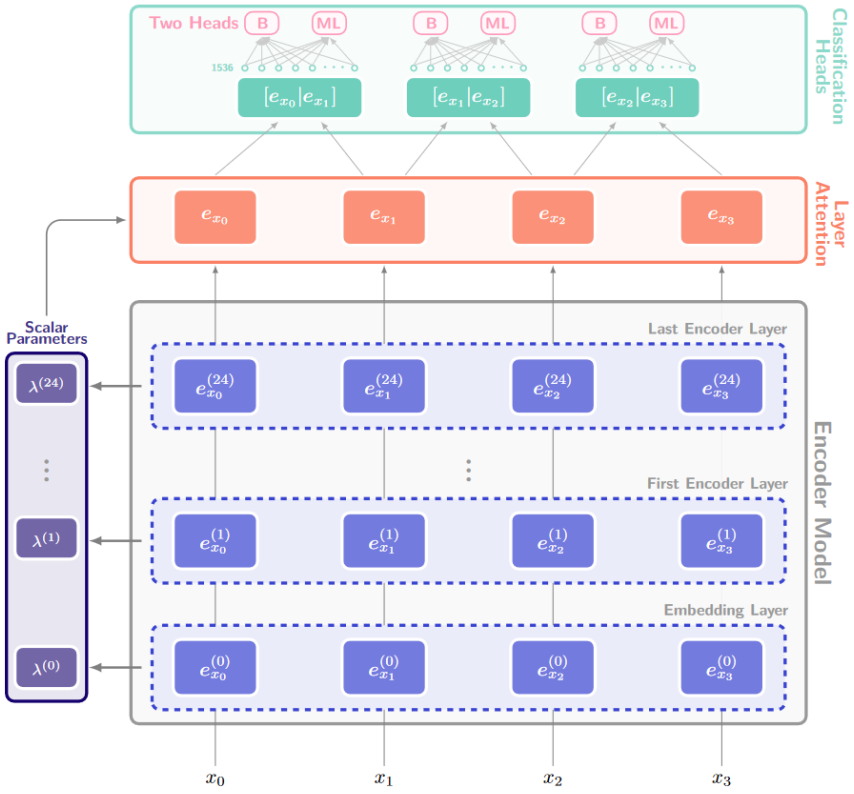

Figure 3: Model architecture used to compete on the SEPP-NLG 2021 shared task. This model follows the architecture proposed by Guerreiro et al. (2021), but with a classification head that simultaneously predicts sentence ends (binary classification) and punctuation marks (multinomial classification).

some of the sentences, mostly consisting of lists of numbers, may contain up to about 200 commas.

4 System Description

As it was previously mentioned, our system architecture extends the architecture proposed by (Guerreiro et al., 2021) which has shown promising, results in multilingual punctuation prediction and capitalization (Rei et al., 2020). This architecture is composed of 3 modules: an Encoder Model, a Layer-wise Attention Mechanism, and a Classification Head. In our experiments to the shared task we replaced the XLM-R base with XLM-R large (Conneau et al., 2020) and also added a new binary classification head for subtask 1 (full-stop prediction).

With that said, when our system receives a document, that document is tokenized using XLM-R tokenizer and divided into several input sequences $\boldsymbol{x^i} = [x_0^i, x_1^i, \ldots, x_{511}^i]$ with 512 sub-words. Then for each input sequence, the encoder will produce an embedding $e_{x_j^i}^{(\ell)}$ for each sub-word x_j^i and each layer $\ell \in \{0, 1, ..., 24\}$. To encapsulate information from all transformer layers into a single embedding, $\boldsymbol{e}_{x_j^i}$, the following layer-wise attention mechanism is used:

$$\boldsymbol{e}_{x_j^i} = \gamma \boldsymbol{E}_{x_j^i}^{\top} \boldsymbol{\Lambda} \quad (1)$$

where γ is a trainable scaling factor, $\boldsymbol{E}_{x_j^i} = [e_{x_j^i}^{(0)}, e_{x_j^i}^{(1)}, \ldots e_{x_j^i}^{(24)}]$ corresponds to the vector of layer embeddings for sub-word x_j^i, and $\boldsymbol{\Lambda} = \mathsf{softmax}([\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(24)}])$ is a vector constituted by the layer scalar trainable parameters which are shared for every sub-word. Finally, we concatenate the embeddings of consecutive words[1] in the input sequence $\boldsymbol{x^i}$ and use those as features for our punctuation (ML – multi-label) and full-stop (B – binary) classification heads. Figure 3 illustrates the described architecture.

5 Experiments

We started our experiments with the exact same hyper-parameters used by Guerreiro et al. (2021). To achieve better performance we also ran an hyper-parameter search using OPTUNA (Akiba et al.,

[1] When a word is divided into several sub-words we use the embedding of the first sub-word to represent the entire word.

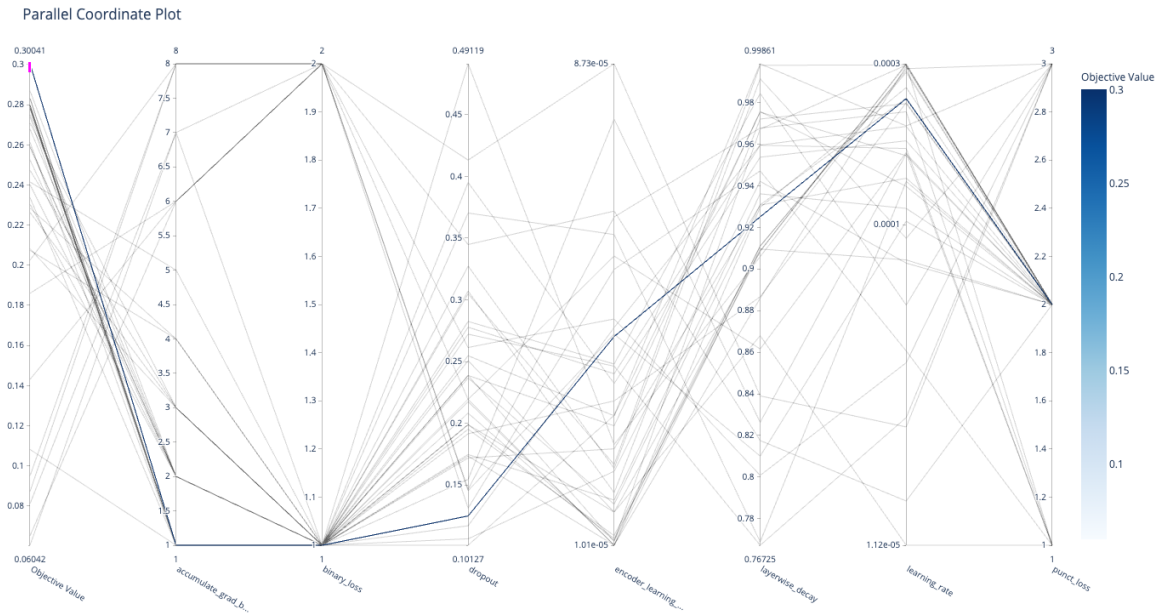

Figure 4: Best trial hyper-parameters highlighted in the OPTUNA search space.

2019). In this section we will describe the training setup and the evaluation metrics used for these experiments.

5.1 Evaluation Setup

The official shared task metric for full-stop prediction is the F1 score of the positive class (sentence end). For the punctuation prediction sub-task the official metric is Macro-F1. Since our developed model performs both tasks at the same time, we also combine those two metrics by multiplying them. Following Guerreiro et al. (2021), we additionally measure the punctuation Slot Error Rate (SER) (Makhoul et al., 2005), a commonly used metric for the task at hand. Also, we discard the "O" (no punctuation) label for calculation of our Macro-F1 scores.

5.2 Training Setup

Our model uses a discriminative fine-tuning strategy with gradual unfreezing by splitting the model parameters into two groups: the XLM-R parameters and the classification heads on top. The encoder parameters are frozen for 0.1% steps of the first epoch. This allows the parameters of the classification heads to adjust to the task objective before changing the pre-trained ones. Then, the entire model parameters are fine-tuned, except the embedding layer that is kept frozen. Keeping the embedding layer frozen allows us to save some GPU memory and fit the entire model into a single 12GB memory GPU.

Evaluation is performed after each epoch using only 50% of the entire development data. The training is interrupted after 2 epochs without improvements on the punctuation task Macro-F1.

5.3 Hyper-parameter Search

We used OPTUNA (Akiba et al., 2019) to search for the optimal hyper-parameters for our model. Our search space was defined as follows:

- Accumulate gradients for 1 to 32 batches (this simulates bigger batches while avoiding memory issues);
- Classification heads dropout between 0.1 and 0.5 with sampling from a uniform distribution;
- Layer-wise learning rate decay between 0.75 and 1.0 with sampling from a uniform distribution;
- Encoder model learning rate between 1e-05 and 1e-04 with sampling from a log-uniform distribution;
- Classification heads learning rate between 1e-05 and 3e-04 with sampling from a log-uniform distribution;
- Full-stop prediction loss with two possible values: 1 and 2;

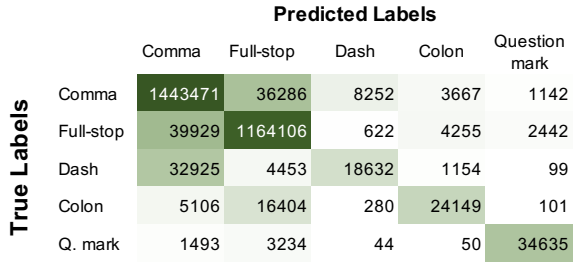

Figure 5: Confusion Matrix for punctuation prediction.

- Punctuation prediction loss with three possible values: 1, 2 and 3.

To speed up the hyper-parameter search we used only 50% of the available training data while keeping the 50% development data described above.

Table 2 reports the results of our baseline against the large models with default hyper-parameters and the best trial results from OPTUNA. As expected, from our table, we can observe that the biggest improvement comes from using XLM-R large in replacement of the base model. We can also observe that further hyper-parameter tuning helps especially in terms of the SER.

Figure 4 shows that the best results were achieved by keeping the encoder_learning_rate low with a high layerwise_decay (above 0.9). The learning_rate for the classification heads is almost 10× higher than the encoder_learning_rate. Finally, the weight of the punctuation prediction loss is set to 2× the weight of the binary prediction loss. Table 1 describes the hyper-parameters used in our baseline along with our final submission.

Hyper-parameter	Baseline	Final submission
Encoder Model	XLM-R (base)	XLM-R (large)
Optimizer	AdamW	AdamW
nº frozen epochs	0.1	0.1
Learning rate	5e-05	2.37e-04
Encoder Learning Rate	3e-05	2.57e-05
Layerwise Decay	1.0	0.925
Batch size	12	8
Loss function	Cross-Entropy	Cross-Entropy
Binary Loss Weight	1	1
Punctuation Loss Weight	1	2
Dropout	0.1	0.125
FP precision	32	16

Table 1: Hyper-parameters used in our final submission compared with the baseline hyper-parameters from Guerreiro et al. (2021).

6 Results

Table 2 shows that, as expected, using a larger encoder improves our results. Also, by using OPTUNA, we were able to further improve our results which means that the models presented by Guerreiro et al. (2021) are under-tuned and could be further improved with a better selection of hyper-parameters.

Looking into the results for individual punctuation marks we can observe that our final submission has a high F1 for *commas*, *full stops* and *question marks*, 96%, 94% and 89% respectively. Yet, the model seems to struggle at predicting *dashes* and *colons* (63% and 39% F1 respectively). By looking at Figure 5, we can observe that, as expected, *dashes* and *colons* are frequently confused with *commas* and *full stops*, respectively. These marks can often be interchanged without loss of meaning. This is further evidence to support the rationale of some proposed approaches to solve this task (Tilk and Alumäe, 2015; Che et al., 2016; Guerreiro et al., 2021), in which *dashes* and *colons* tend to be aggregated with the *commas* and *full stops* labels, respectively.

7 Conclusions and future work

We have described a multilingual model that is able to simultaneously detect sentence boundaries, and to predict 5 different punctuation marks over 4 different languages (English, German, French and Italian). The model was adapted from (Guerreiro et al., 2021), and used by the Unbabel & INESC-ID team for the 1st Shared Task on Sentence End and Punctuation Prediction in NLG Text (SEPP-NLG 2021), achieving one of the top results. The results confirm that it is possible to achieve state-of-the-art results using a single multilingual model for both tasks and multiple languages. This result supports what was already observed in the experiments performed by (Guerreiro et al., 2021). The code used to produce the results is publicly available at: https://github.com/Unbabel/caption/tree/shared-task.

In the future, we plan to extend this work to include other language families, such as Semitic and Slavic languages. Moreover, we would like to extend our setup to be capable of simultaneously solving the capitalization task too. Having one single multilingual model that is capable of identifying sentence boundaries, punctuation marks and proper capitalization would constitute a major step

Development Models	SER↓	Binary F1↑	Macro F1↑	Macro x Binary↑
Baseline (Guerreiro et al., 2021)	0.265	0.926	0.399	0.369
XLM-R large (default)	0.243	0.944	0.411	0.388
XLM-R large OPTUNA	0.214	0.944	0.444	0.419

Table 2: Results of our models on the shared task development data. Our baseline model is trained with the exact same setup as the multilingual models from Guerreiro et al. (2021). Then we decided to replace XLM-R base by XLM-R large. Finally to further improve our results we used OPTUNA to search over the hyper-parameters space described in Section 5.3. Note that these experiments were performed using the shared task corpus V1.

towards recovering from ASR recognition errors and translation errors from MT systems.

Acknowledgments

This work was supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020 and by the P2020 Program through projects "Unbabel Scribe" and "MAIA" supervised by ANI under contract numbers 038510 and 045909 respectively.

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 2623–2631, New York, NY, USA. Association for Computing Machinery.

F. Batista, D. Caseiro, N. Mamede, and I. Trancoso. 2008. Recovering capitalization and punctuation marks for automatic speech recognition: Case study for portuguese broadcast news. *Speech Commun.*, 50(10):847–862.

Fernando Batista, Diamantino Caseiro, Nuno J. Mamede, and Isabel Trancoso. 2007. Recovering punctuation marks for automatic speech recognition. In *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27-31, 2007*, pages 2153–2156. ISCA.

Fernando Batista, Helena Moniz, Isabel Trancoso, and Nuno J. Mamede. 2012. Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts. *IEEE Transactions on Audio, Speech and Language Processing, Special Issue on New Frontiers in Rich Transcription*, 20(2):474–485.

Fernando Batista, Helena Moniz, Isabel Trancoso, Hugo Meinedo, Ana Isabel Mata, and Nuno J. Mamede. 2010. Extending the punctuation module for european portuguese. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, pages 1509–1512. ISCA.

Fernando Batista, Isabel Trancoso, and Nuno J Mamede. 2009. Comparing automatic rich transcription for portuguese, spanish and english broadcast news. In *Automatic Speech Recognition and Understanding, 2009. ASRU 2009. IEEE Workshop on*, pages 540–545. IEEE.

Doug Beeferman, Adam Berger, and John Lafferty. 1998. Cyberpunc: a lightweight punctuation annotation system for speech. *ICASSP*, pages 689–692.

Y. Cai and D. Wang. 2019. Question mark prediction by BERT. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 363–367.

Roldano Cattoni, Nicola Bertoldi, and Marcello Federico. 2007. Punctuating confusion networks for speech translation. In *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27-31, 2007*, pages 2453–2456. ISCA.

Xiaoyin Che, Cheng Wang, Haojin Yang, and Christoph Meinel. 2016. Punctuation prediction for unsegmented transcript based on word vector. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 654–658, Portorož, Slovenia. European Language Resources Association (ELRA).

H. Christensen, Y. Gotoh, and S. Renals. 2001. Punctuation annotation using statistical prosody models. In *Proc. of the ISCA Workshop on Prosody in Speech Recognition and Understanding*, pages 35–40.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Markus Freitag, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan, and Wolfgang Macherey. 2021. Experts, errors, and context: A large-scale study of human evaluation for machine translation.

Nuno Miguel Guerreiro, Ricardo Rei, and Fernando Batista. 2021. Towards better subtitles: A multilingual approach for punctuation restoration of speech transcripts. *Expert Systems With Applications (under review)*.

Mary Harper, Bonnie Dorr, John Hale, Brian Roark, Ishak Shafran, Matthew Lease, Yang Liu, Matthew Snover, Lisa Yung, Anna Krasnyanskaya, and Robin Stewart. 2005. Parsing and spoken structural event detection. In *2005 Johns Hopkins Summer Workshop Final Report*.

Jing Huang and Geoffrey Zweig. 2002. Maximum entropy modeling for punctuation from speech. In *Proceedings of ICSLP*.

D. Jones, E. Gibson, W. Shen, N. Granoien, M. Herzog, D. Reynolds, and C. Weinstein. 2005. Measuring human readability of machine generated text: three case studies in speech recognition and machine translation. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, volume 5, pages v/1009–v/1012.

J. Kim and P. C. Woodland. 2001. The use of prosody in a combined system for punctuation generation and speech recognition. In *Proc. of Eurospeech*, pages 2757–2760.

Ji-Hwan Kim and Philip C. Woodland. 2003. A combined punctuation generation and speech recognition system and its performance enhancement using prosody. *Speech Communication*, 41(4):563 – 577.

Seokhwan Kim. 2019. Deep recurrent neural networks with layer-wise multi-head attentions for punctuation restoration. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7280–7284.

Ondrej Klejch, Peter Bell, and Steve Renals. 2017. Sequence-to-sequence models for punctuated transcription combining lexical and acoustic features. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5700–5704.

Junwei Liao, Sefik Emre Eskimez, Liyang Lu, Yu Shi, Ming Gong, Linjun Shou, Hong Qu, and Michael Zeng. 2020. Improving readability for automatic speech recognition transcription.

Yang Liu, Elizabeth Shriberg, Andreas Stolcke, Dustin Hillard, Mari Ostendorf, and Mary Harper. 2006. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transaction on Audio, Speech and Language Processing*, 14(5):1526–1540.

Yang Liu, Elizabeth Shriberg, Andreas Stolcke, Barbara Peskin, Jeremy Ang, Dustin Hillard, Mari Ostendorf, Marcus Tomalin, Phil Woodland, and Mary Harper. 2005. Structural metadata research in the EARS program. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, Philadelphia, USA.

Wei Lu and Hwee Tou Ng. 2010. Better punctuation prediction with dynamic conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 177–186, Cambridge, MA. Association for Computational Linguistics.

K. Makhija, T. Ho, and E. Chng. 2019. Transfer learning for punctuation prediction. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 268–273.

J. Makhoul, A. Baron, I. Bulyko, L. Nguyen, L. Ramshaw, D. Stallard, R. Schwartz, and B. Xiang. 2005. The effects of speech recognition and punctuation on information extraction. In *INTERSPEECH-05*, pages 57–60.

Evgeny Matusov, Arne Mauser, and Hermann Ney. 2006. Automatic sentence segmentation and punctuation prediction for spoken language translation. In *International Workshop on Spoken Language Translation*, pages 158–165, Kyoto, Japan.

Joanna Mrozinsk, Edward WD Whittaker, Pierre Chatain, and Sadaoki Furui. 2006. Automatic sentence segmentation of speech for automatic summarization. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*.

M. Ostendorf, E. Shriberg, and A. Stolcke. 2005. Human language technology: Opportunities and challenges. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, Philadelphia.

Mari Ostendorf, Benoit Favre, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Dustin Hillard, Julia Hirschberg, Heng Ji, Jeremy G. Kahn, Yang Liu, Sameer Maskey, Evgeny Matusov, Hermann Ney, Andrew Rosenberg, Elizabeth Shriberg, Wen Wang, and Chuck Wooters. 2008. Speech segmentation and spoken document processing. *IEEE Signal Processing Magazine*, 25(3):59–69.

Stephan Peitz, Markus Freitag, Arne Mauser, and Hermann Ney. 2011. Modeling punctuation prediction as machine translation. In *International Workshop on Spoken Language Translation*, pages 238–245, San Francisco, CA, USA.

Ricardo Rei, Nuno Miguel Guerreiro, and Fernando Batista. 2020. Automatic truecasing of video subtitles using bert: A multilingual adaptable approach. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 708–721, Cham. Springer International Publishing.

Elisabeth Shriberg. 2005. Spontaneous speech: How people really talk, and why engineers should care. In *Proc. of Eurospeech - 9^{th} European Conference on Speech Communication and Technology (Interspeech 2005)*, pages 1781 – 1784, Lisbon, Portugal.

Ottokar Tilk and Tanel Alumäe. 2015. LSTM for punctuation restoration in speech transcripts. In *INTERSPEECH*.

Ottokar Tilk and Tanel Alumäe. 2016. Bidirectional recurrent neural network with attention mechanism for punctuation restoration. In *INTERSPEECH*, pages 3047–3051.

Nicola Ueffing, Maximilian Bisani, and Paul Vozila. 2013. Improved models for automatic punctuation prediction for spoken and written text. In *INTERSPEECH*.

Jiangyan Yi and Jianhua Tao. 2019. Self-attention based model for punctuation prediction using word and speech embeddings. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7270–7274.

Klaus Zechner. 2002. Automatic summarization of open-domain multiparty dialogues in diverse genres. *Computational Linguistics*, 28(4):447–485.