ETC-NLG: End-to-end Topic-Conditioned Natural Language Generation

Ginevra Carbone[1] and Gabriele Sarti[1,2]
[1] Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy
[2] International School for Advanced Studies (SISSA), Trieste, Italy
ginevra.carbone@phd.units.it, gsarti@sissa.it

Abstract. Plug-and-play language models (PPLMs) enable topic-conditioned natural language generation by combining large pre-trained generators with attribute models to steer the predicted token distribution towards selected topics. Despite their efficiency, the large amounts of labeled texts required by PPLMs to effectively balance generation fluency and proper conditioning make them unsuitable to low-resource scenarios. We present **ETC-NLG**, an approach leveraging topic modeling annotations to produce **E**nd-to-end **T**opic-**C**onditioned **N**atural **L**anguage **G**enerations over emergent topics in unlabeled document collections. We test our method's effectiveness in a low-resource setting for Italian and perform a comparative evaluation of ETC-NLG for Italian and English using a parallel corpus. Finally, we propose an evaluation method to automatically estimate the conditioning effectiveness from generated utterances.

Keywords: Natural language generation · Topic modeling · Conditioned generation · Neural language models

1 Introduction

Pre-trained neural language models can be used for natural language generation (NLG) by autoregressively sampling the most probable token from the learned vocabulary distribution given the previous context. Among the most effective autoregressive models, GPT variants [15,16,4] follow the two-step process originally introduced by ULMFiT [9], combining an unsupervised pretraining over massive textual resources with a task-specific fine-tuning to solve a variety of different problems. Despite their effectiveness for standard NLG, GPT-like models are still mostly inefficient for advanced forms of NLG, such as topic and sentiment-conditioned generation, requiring fine-tuning with attribute-specific data or even radically changing the model's architecture [10] to allow for better control over generated outputs. *Plug-and-play language models (PPLMs)* [5] were recently introduced to counter this tendency, allowing users to efficiently generate controlled text by combining a standard pre-trained language model generator with a discriminator network that learns to differentiate attributes and to steer generation towards the selected conditioning.

While simple BoW models have been demonstrated to perform effective conditioning in the context of sentiment analysis [12], differentiating abstract thematic categories often requires training more sophisticated discriminators able to capture semantic relations between entities from distributed representations. For this reason, PPLMs need large quantities of annotated documents to train discriminators capable of successfully steering the generation process. This fact makes them mostly unsuitable for low-resource domains and languages where such annotations are often unavailable. To address this weakness, we propose **ETC-NLG**, an approach leveraging the efficiency of PPLMs and the effectiveness of contextual [2] and combined [1] topic models to enable an **E**nd-to-end **T**opic-**C**onditioned **N**atural **L**anguage **G**eneration from unlabeled document collections.[1] We begin by evaluating the effectiveness of ETC-NLG on the Svevo Corpus [7,17], a topic-annotated Italian epistolary corpus containing archaic and dialectal terms. We then compare the effectiveness of Italian neural language models used by ETC-NLG with their widely-used English counterparts, testing their topic modeling and conditioned generation performances on a portion of the EuroParl Italian-English parallel corpus [11]. We finally assess the quality of generated utterances and propose an automatic method to evaluate their conformity to the selected conditioning topic.

2 Background

Combined and Contextual Topic Models Combined topic models [1] extend the Neural-ProdLDA [18] variational approach by concatenating the input BoW document representation with pre-trained contextual embeddings produced by a neural language model. An inference network maps the resulting vectors to a latent representation, which is then variationally sampled by a second decoder network to reconstruct the document BoW, effectively approximating the standard Dirichlet prior [3] with normally-distributed samples. Fully-contextual topic models [2] follow the same principle, but language models' contextual embeddings entirely replace the BoW representation. The main advantage of these approaches over classical LDA-like methods for topic modeling is the use of semantically-informed representations, which were shown to improve intra-topic coherence and provide more meaningful classifications.

Plug-and-play Language Models Plug-and-play language models [5] flexibly combine large pre-trained language models with cheap discriminators, to perform controlled text generation. These models follow a Bayesian approach by including:

- An *unconditional language model* generator, acting as a prior probability distribution $p(x)$ over text.
- An *attribute model* discriminator $p(a|x)$, expressing the likelihood of an attribute a given text x.
- The resulting *conditional language model* $p(x|a)$ used for conditional autoregressive generation.

[1] Code available at `https://github.com/gsarti/ETC-NLG`.

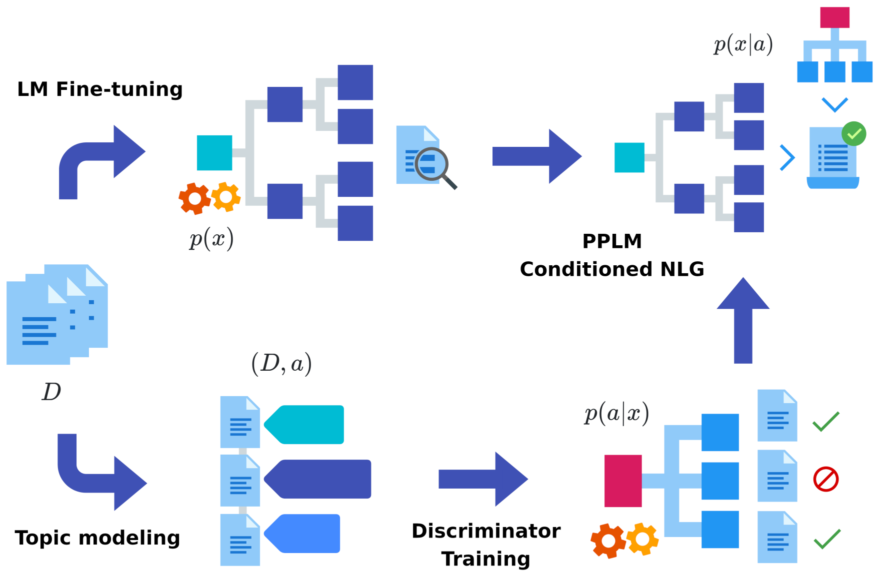

Fig. 1: A visual representation of the End-to-end Topic-Conditioned Natural Language Generation (ETC-NLG) pipeline. An unlabeled document collection is used to train a language model (top) and a topic model (bottom). Automatic topic annotations are used to condition the generation of a PPLM through a discriminator network.

This approach is especially appealing since tiny and cheap attribute models can produce excellent conditional generative models when combined with powerful pre-trained architectures such as GPT-2 [16]. Performances resulting from those models were shown to be comparable to those of cumbersome fully-conditional language models like CTRL [10] with far less computation. The objective of PPLM is that of steering the representation of text toward high log-likelihood values for both the conditional attribute model $p(a|x)$ and the unconditional language model $p(x)$. Maximizing $\log p(a|x)$ ensures that x will more likely possess attribute a, while maximizing $p(x)$ guarantees fluency within the generated text.

3 Methodology

Figure 1 presents a visual representation of our approach that builds upon the PPLM architecture. We start from an unlabeled document collection D (left) and fine-tune a neural language model generator to adapt its predicted distribution over the vocabulary to the current setting, obtaining our unconditional language model $p(x)$ (top-left). We then perform automatic topic modeling over D using either combined or contextual topic models. Trained topic models are subsequently used to annotate each document in D with its most relevant topic, obtaining a collection of topic-annotated documents (D, a) (bottom-left). Automatic topic annotations are used to train an attribute model discriminator $p(a|x)$ that predicts document topics given their contextual embedding representations (bottom-right). Finally, we merge the two networks to obtain a PPLM conditional language model $p(x|a)$ for topic-conditioned utterances (top-right). Language and topic modeling steps can be performed in parallel to optimize training times by taking advantage of the independence between the generative and discriminative components of ETC-NLG.

While this approach is particularly useful when dealing with insufficient labeled data and low-resource scenarios, since topic labels are inferred, the production of meaningful sentences heavily relies on topic modeling quality. We discuss this perspective after the experimental results of Section 4.

4 Experimental Results

Our experimental objectives are three-fold: first, we test ETC-NLG on the Italian subset of the epistolary corpus of Italo Svevo [7], a famous Italian author of the early 20th century, to quantify the impact of dialectal and archaic expressions on the quality of generated sentences. Secondly, we compare the performances of ETC-NLG on Italian and English by leveraging a portion of the European Parliament Proceedings (EuroParl) parallel corpus [11]. Finally, we perform an empirical evaluation of the obtained results and present an intrinsic evaluation method based on topic modeling quality over conditionally-generated texts.

Data The Svevo Corpus contains 5419 sequences ranging from few words to multiple sentences. Each sequence is annotated with one or more topics by two domain experts using a set of five main topics (family, literature, work, travel, health) and five subtopics that were found during a previous analysis of the corpus [17]. We aggregate most of the sub-topics with main topics to reduce sparsity and obtain a final set of documents annotated by 6 topics: *family* (1669 seq.), *wife* (1298 seq.), *travel* (821 seq.), *health* (552 seq.), *literature* (544 seq.) and *work* (535 seq.). The EuroParl corpus does not contain topic annotations and comprises almost 2 million parallel sentences collected from parallel Italian-English proceedings of the European Parliament. We only select the first 50'000 sentences for our modeling experiments.

4.1 Topic Modeling

We test both combined and contextual topic modeling approaches using RoBERTa [13], a widely-known improvement over the BERT encoder [6], and UmBERTo [8], a RoBERTa-based encoding language model trained on crawled Italian web data, producing respectively English and Italian contextual representations. We leverage the base variants of both models available through the HuggingFace Transformers framework [19]. Contextual embeddings are sampled either alone or alongside bag-of-words representations in a variational framework to improve topic coherence [2].

[2] See works by Bianchi et al. [2,1] for specifications on models and metrics.

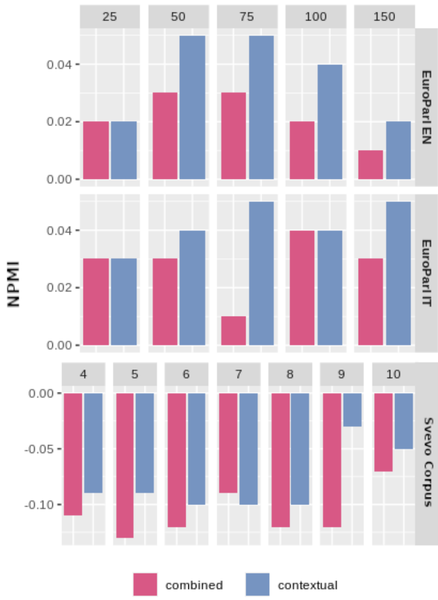

Fig. 2: NPMI scores for contextual and combined topic models over the three corpora with variable topic counts. Higher scores correspond to higher relatedness between topic words.

	Contextual			Combined		
	α	ρ	τ	α	ρ	τ
Svevo Corpus-4	.93	.98	-.09	.99	1.00	-.11
Svevo Corpus-5	.85	.97	-.09	.97	1.00	-.13
Svevo Corpus-6	.72	.92	-.10	.91	.98	-.12
Svevo Corpus-7	.71	.94	-.10	.91	1.00	-.09
Svevo Corpus-8	.73	.93	-.10	.79	.95	-.12
Svevo Corpus-9	.69	.94	-.03	.77	.96	-.12
Svevo Corpus-10	.64	.91	-.05	.78	.96	-.07
EuroParl-IT-25	.79	1.00	.03	.67	.99	.03
EuroParl-IT-50	.54	.99	.04	.41	.98	.03
EuroParl-IT-75	.42	.99	.05	.24	.96	.01
EuroParl-IT-100	.33	.98	.04	.23	.96	.04
EuroParl-IT-150	.23	.98	.05	.17	.96	.03
EuroParl-EN-25	.82	1.00	.02	.72	.99	.02
EuroParl-EN-50	.56	.99	.05	.44	.99	.03
EuroParl-EN-75	.43	.99	.05	.32	.98	.03
EuroParl-EN-100	.35	.99	.04	.15	.94	.02
EuroParl-EN-150	.16	.96	.02	.13	.94	.01

Table 1: Result of topic modeling evaluation using topic diversity (α), inverted RBO (ρ) and NPMI (τ) metrics. **Bold** models were selected for conditioned generation experiments.

Given the different sizes of tested corpora, we evaluate combined and contextual models' performances by varying the number of topics between 3 and 10 for the Svevo corpus and between 25 and 150 for the EuroParl corpora. We use three metrics capturing topic coherence and diversity: i) Normalized Pointwise Mutual Information (NPMI, τ), measuring the relatedness of top-10 topic words given the empirical frequency of corpus words; ii) Topic Coherence (α), i.e., the average of pairwise similarities between word embeddings of top-25 topic words across topics for a semantic perspective to topic coherence; and iii) Rank-Biased Overlap diversity (Inverted RBO, ρ), a measure of disjointedness between topics weighted on word rankings. Figure 2 reports the NPMI scores for combined and contextual models over topic ranges for the three corpora. Topics generated by the contextual model are generally more coherent across all topic counts, with EuroParl topics being more coherent than those of the Svevo corpus for both Italian and English languages.

Table 1 reports the whole set of evaluated metrics for all models on all corpora. From the results, we can see a clear tradeoff between the number of selected topics and the quality of α and ρ metrics, while NPMI does not appear to be affected by topic counts. For our generation experiments, we choose the 6-topics contextual model for

Test performances	Svevo Corpus	EuroParl IT	EuroParl EN
Fine-tuned LM perplexity	37.00	14.04	6.80
Discriminator test accuracy (Gold)	62%	-	-
Discriminator test accuracy (Contextual)	51%	95%	91%

Table 2: A contextual topic model is used to produce labels for training the contextual discriminator on each corpus. No gold labels are available for the EuroParl corpus.

Base LM	GPT-2 (EN) [16] GePpeTto (IT) [14]
LM fine-tuning	epochs = 2 max sequence length = 128
Discriminator	epochs = 10 max sequence length = 128
PPLM	output sequences length = 60 iterations = 15 repetition penalty = 1.5 window length = 0 horizon length = 5 top-k = 10 step size = 0.05 $0.9 \leq$ gm scale ≤ 0.99 $1. \leq$ temperature ≤ 1.5

Table 3: Training hyperparameters were optimized over Svevo Corpus and reused for EuroParl.

the Svevo corpus to match the number of gold topic labels empirically set by human annotators, and the 75-topic contextual models for both Italian and English EuroParl, given their strong performances in both topic coherence and NPMI. In the upcoming sections, it should be assumed that those are the only models used to produce annotations for training the discriminators unless otherwise mentioned.

4.2 Conditioned Text Generation

We use GPT-2 as the PPLM generator for the English EuroParl dataset and its Italian counterpart GePpeTto [14] for Svevo and Italian EuroParl datasets. In all such cases, we observe that 2 fine-tuning epochs are enough to obtain adequate models with the lowest perplexity values (Table 2). We also observe that using a low number of iterations for LM fine-tuning (2 to 5) is often optimal. Given the well-known computational limitations of Transformers, due to the $O(n^2)$ self-attention computation performed at each layer, sentences from training corpora were cut at the last punctuation symbol before a fixed maximum sequence length. We chose a maximum of 128 tokens for our experiments, meaning that LMs were trained on sentences with a variable number of tokens up to this value.

A lightweight transformer encoder followed by a dense classification head is used as the discriminator and trained on gold labels for Svevo letters and topic modeling labels produced by the selected topic models for all corpora. Examples of Svevo gold labels are reported in Table 6 of the Appendix. For the EuroParl corpus, the discriminator was trained by selecting the ten most frequent topics from automatic annotations. Table 2 shows the best performances obtained in all scenarios.

[Wife] **La tua** assenza mi secca. Non ho nulla da dirti e spero che tu potrai sapere se la lettera sarà spedita in qualche città della mia vita o meno, a Venezia oppure a Milano!
[Travel] **Un giorno** mi disse che per me sarebbe stato meglio di non venirci a prendere. Se ci fossero stati due o quattro giorni sarei partito senza di loro e avrei fatto un viaggio simile, in una città più bella della stessa Parigi dove il sole si leva.
[Literature] **Un giorno** ti scriverò. Non ho scritto che il primo bacio sia stato quello di Olga. Ho ricevuto la lettera di Letizia e Marco, i due francesi, con le lettere d'ieri.
[Work] **Se potessi** fare un simile sacrificio di tempo in ufficio, sarei molto meglio esposto a questo rischio; se tu mi dici di aver bisogno d'operazioni (per esempio la posta) io direi che il lavoro è più facile ma bisogna farlo bene per avere delle idee nuove.
[Health] **Se potessi** avere un po' di riposo per la mia giornata, avrei fatto una grande visita al mio medico di famiglia. Mi disse: ≪Sai come ti scrivo e mi dici quale cura è quella del tuo male≫. Poi rispose che non sarebbe stata necessaria l'≪iniezione dl≫ ma solo una o più visite (che può portare allungarsi fino a 2 settimane).
[Family] **La tua** lettera non ti scriverà niente. Tanti saluti da Paola, mia cara amica e di tutta la famiglia Villa Veneziani a Trieste; una grande giornata per tutti!

Table 4: Examples of ETC-NLG topic-conditioned generation from the Svevo corpus using gold labels, with `temperature` values between $[1, 1.5]$ and `gm_scale` between $[0.9, 0.99]$. Blue text is the conditioning topic, **bold** represents prefix context.

The best topic model trained on Svevo letters brought the discriminator to maximum test accuracy of 51%, under optimal training settings among those that were tested. In terms of test accuracy, we observe that a poor discriminator represents a significant bottleneck in generating high-quality sentences from the contextual PPLM. The discriminator trained on gold labels achieves higher test accuracy (Table 2), showing that manually-annotated sequences are also more separable from a contextual embedding perspective. From Table 2, we answer our first two experimental objectives by noticing that:

– The dialectal and archaic nature of the Svevo corpus severely cripples both generator and discriminator performances, making conditioned NLG more challenging.

– Both the generator and discriminator obtain comparable scores on the EuroParl parallel corpora, suggesting that the quality of pre-trained contextual representations of language models, both the one used for language modeling and the one used for topic modeling, is similar across the two languages.

We note that two fundamental components for achieving good PPLM performances are a language model with low perplexity over the selected corpus and a discriminator model with high test accuracy on corpus annotations. The combination of these two properties guarantees a generation of text that is both fluent and adequately contextualized.

4.3 Evaluation

After fine-tuning the language model and training the discriminator, we can produce conditioned sentences according to the PPLM scheme. We choose four different neutral prefix sentences (see Table 7) for each model and generate three conditioned sentences for all possible combinations of topics and prefixes. We produce a total of 72 sentences on the Svevo Corpus (plus other 72 on gold labels for human evaluation) and 120 sentences each for both EuroParl-IT and EuroParl-EN.

Human evaluation on gold labels We start our assessment by manually estimating the quality of conditioned generations. We wish to emphasize that our human evaluation is solely intended as a qualitative assessment performed without external participants' help and is not supported by any statistically significant evaluation. Sentences generated by the PPLM model based on Svevo gold labels show some weaknesses in performing proper conditioning, as expected after the discriminator's poor results, but were generally well-formed and coherent from both a morphological and a syntactic perspective. We experiment by setting higher values for the `step_size` parameter, controlling the size of a gradient step, the `temperature` parameter, inducing a decreased model confidence in its top predictions, and the `gm_scale` parameter, which accounts for the weight of perturbed probabilities during the history updates, to produce a model with stronger conditioning. Values chosen for these hyperparameters are reported in Table 3. Examples of conditioned generation on the Svevo Corpus presented in Table 4 shows how ETC-NLG can produce meaningful sentences despite the relatively high perplexity achieved by GePpeTto generator on the epistolary corpus. Additional examples of generated sentences for all corpora are provided in Table 5.

4.4 Automated evaluation from topic models

We conclude our analysis by proposing a method to assess the conditioning intensity achieved in text generation automatically. We use the same contextual topic models that were originally used to label the corpora in the ETC-NLG pipeline and leverage them to predict each conditionally-generated sentence's most likely topic. Then, we judge the quality of topic conditioning by looking at the resulting confusion matrices between actual and predicted conditioning topics. We point out that this method can only evaluate the consistency between the generated sentences and the prior topic model but cannot estimate their conditioning quality in an absolute sense. Fig. 3 shows the results obtained from strong conditioning on EuroParl-IT and EuroParl-EN while preserving high fluency in both cases. We can observe from the results that, by employing proper parametrization, it is possible to generate sentences visibly conditioned using the ETC-NLG pipeline. Our parameter search experiments, which were not included for brevity, suggest that hyper-parameter tuning significantly influences the system's ability to generate conditioned text reliably. In particular, stronger conditioning is achieved by lowering `step_size` and `gm_scale` values, while the `temperature` parameter appears to have a very moderate effect on conditioning. Training hyperparameters are reported in Table 3 for reproducibility purposes. A closer manual inspection of confusion matrices suggests that most misclassifications occur on topics that appear to us as ill-defined, confirming the importance of proper topic modeling for better automatic evaluation.

5 Conclusions

In this work, we presented ETC-NLG, an end-to-end method leveraging topic modeling annotations on unlabeled text corpora to generate topic-conditioned sentences in natural language. We highlighted this technique's strengths and weaknesses for both English and Italian, focusing mainly on the more challenging case of dialectal and archaic language in the Italian setting. We performed a thorough analysis of both generation and topic modeling performances and concluded by presenting an experimental method to automatically evaluate the effectiveness of conditioning in the generated samples. Controlling the context of generated language is crucial for any real-world application involving human-machine interaction. Automatizing the evaluation procedure of generation models has the potential to significantly improve their usability in realistic settings. Our method aims to account for insufficient labeled training data and can be used to produce high quality conditioned text when provided with suitable topic models and a good set of parameters to balance generation fluency and conditioning strength.

Future developments of our approach should focus on two main bottlenecks in the ETC-NLG pipeline: developing better and more robust topic models for the automatic annotation of low-resource text and dealing with the computationally-heavy generation of conditioned sentences.

Svevo Corpus
[senilità–devotissimo–joyce–amicare–carissimo] **La tua** lettera di ieri, la seguente: ≪Se ti dicessi che non sei un critico d'arte o uno scrittore ma una persona capace e forte...≫. Non credo tu ne abbia bisogno. Se mi fosse possibile sarebbe meglio che anche io finisse per scriverti senza avere a cui [senilità–devotissimo–joyce–amicare–carissimo]
[scell–halperson–roncegno–finito–scala] **Se potessi** essere a Salso sarei felice di vedermi ancora più vicino. Sto benone. Ho avuto un'idea della mia assenza ma non ho potuto scriverla in nessun modo perché la mia lettera è perduta per il momento e sono stanco solo che dopo aver letto le tue lettere [decembre–tribel–raffreddore–debole–capanna]
EuroParl EN
[congratulate–excellent–rapporteur–thank–congratulation] **It is** therefore important to make the necessary adjustments, especially in relation human rights and democratic principles. There are a number of points which I believe are crucial, but this also affects us all: it must be equal for all our [racism–xenophobia–violence–minority–ethnic]
[state–member–national–small–large] **You did** not request an additional explanation. If you look at what the Committee on Economic and Monetary Affairs has to say, I do agree with your positions. Most of all is that we must bear in mind the need for the single currency - as it was [market–euro–company–investment–service]
[peace–process–negotiation–agreement–israel] **I would** like first to say that I supported the report. We have begun negotiations with Croatia and Macedonia overall, but we will not be able either to join forces in this area or to continue working on it unless there are new conditions. [peace–process–negotiation–agreement–israel]
EuroParl IT
[acqua–mare–pescare–rifiuto–inquinamento] **In questo** modo, si potrà garantire una migliore protezione dell'ambiente e delle risorse a livello mondiale. Per il momento non sono ancora soddisfacenti le previsioni della commissione per l'industria, la ricerca e lo sviluppo tecnologico sulla riduzione del tenore di zolfo nei combustibili liquidi. [acqua–mare–pescare–rifiuto–inquinamento]
[umano–fondamentale–libertà–diritto–carta] **Si dovrebbe** invece essere più cauti quando si tratta di stabilire un nesso fra la politica della concorrenza e le condizioni sociali. L'idea che l'Unione europea sia uno strumento per il benessere delle sue popolazioni è in realtà una falsa illusione, perché non esiste niente al mondo reale. [umano–fondamentale–libertà–diritto–carta]
[produrre–cioccolato–produttore–consumatore–qualità] **Si dovrebbe** prestare maggiore attenzione alla prevenzione e al ripristino degli habitat naturali. La biodiversità deve costituire uno dei principali problemi di tutte le politiche comunitarie, in quanto è l'unico criterio valido per decidere come affrontare i cambiamenti climatici, soprattutto nel settore agricolo; pertanto dobbiamo tenere conto dell'importanza del [acqua–mare–pescare–rifiuto–inquinamento]

Table 5: Examples of ETC-NLG topic-conditioned generations from all corpora using automatically-produced labels. Blue text represents conditioning topic, **bold text** represents prefix context provided for generation, red text represents topic model prediction during automatic evaluation.

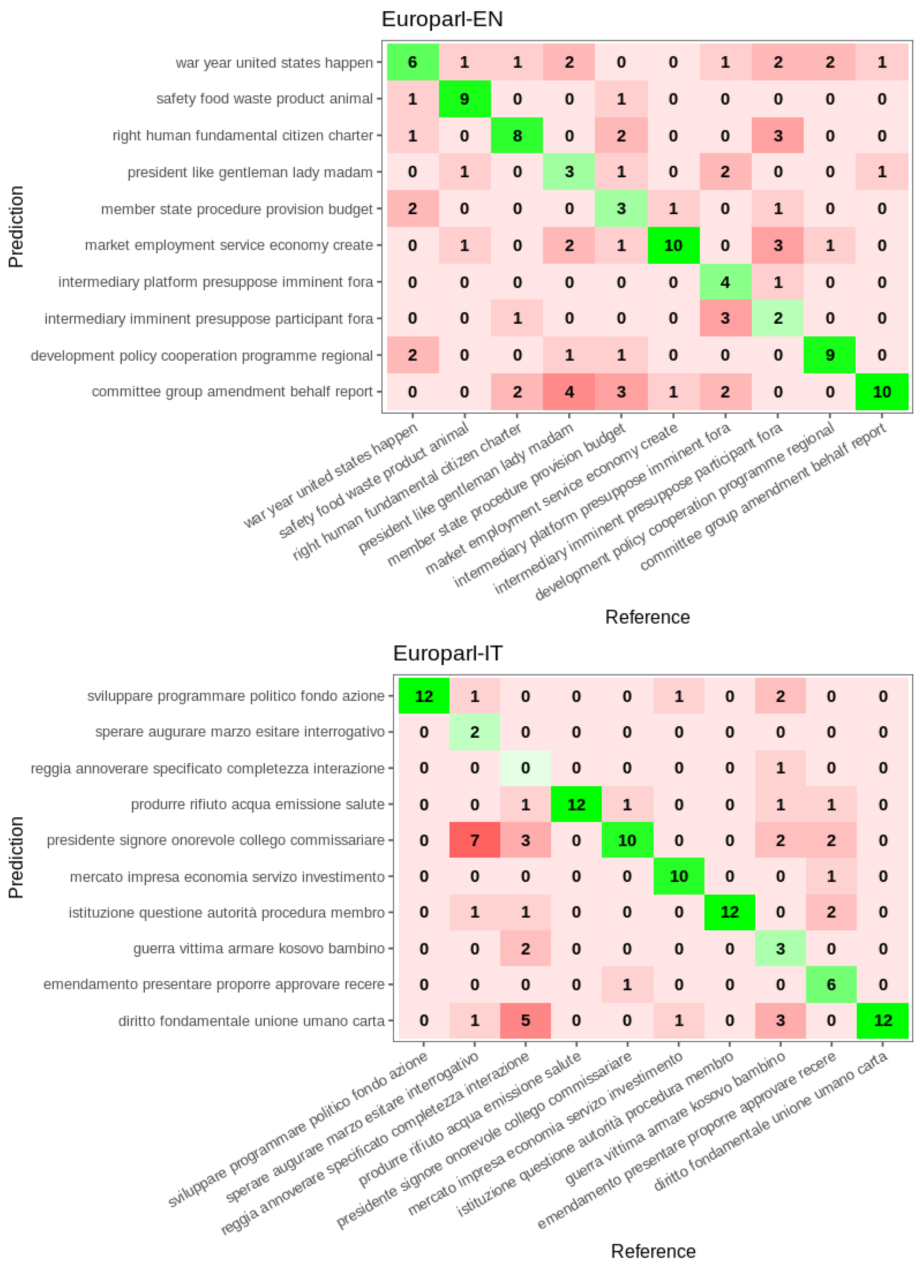

Fig. 3: Confusion matrices obtained by predicting the most likely topic from conditionally-generated sentences produced by ETC-NLG on the EuroParl parallel corpus. We leverage the same contextualized model used to annotate the unlabeled corpora to generate topic predictions and compare those against conditioning labels for sentences generated with strong conditioning parameters (`step_size`$=0.3$, `gm_scale`$=0.95$). Rows represent topic predictions, columns are label references.

Appendix

6 Annotated Svevo Corpus

Table 6 reports examples of Svevo corpus annotations. Sentences can have multiple labels with possibly overlapping meaning, causing a higher uncertainty in predictions for the resulting discriminator model. Only the main topics presented in Section 4 were considered during Svevo gold discriminator training.

Svevo corpus with gold labels
Nel corso di questo mese sarò probabilmente a Parigi. Verrò a salutarla. Porterò con me gli articoli più importanti (pochissimi) che mi furono dedicati. Già non credo ch'Ella abbia premura. Per il momento Ella ha molte altre cose cui pensare prima che all'articolo da dedicare a me. Ha visto l'articolo di Marcel Thiébaut nella ≪ Revue de Paris≫ del 15 Novembre? Per dire il vero l'articolo del ≪Baretti≫ non mi piacque molto. Nell'ultimo ≪Convegno≫ di Milano c'è un articolo di Sergio Solmi pieno di belle osservazioni. [Travel, Literature]
E quando il momento è brutto — ed è anzi il momento brutto quasi sempre perché è l'ora in cui si sentirebbe più il bisogno di una lieta compagnia, di un appoggio, di un incoraggiamento —, scrivendoti cicco quasi sempre e dovresti ancora essere contenta che tante volte il pensiero a te mi mitiga. Iersera presi bromuro e una di quelle tue Purgen e dormii bene. [Wife, Health][Health–Smoking]
Ma io desidero vivamente ch'Ella conosca anche l'originale del Zeno. Non soltanto perché è là cosa che — checché ne dicano i malevoli — scrissi meglio, ma anche perché per volere del Gallimard la traduzione fu falcidiata di non meno di 100 pagine. Come se in francese non esistessero dei romanzi più lunghi del mio. Fui un po' loquace, è vero, ma è doloroso, specialmente per un loquace, di sentirsi tagliare la parola. Una ferita che — secondo il Freud — è esposta alla soppressione patologica. [Literature]

Table 6: Examples from the Svevo Corpus with gold main and sub-topic annotations.

7 Prefix sentences

Svevo	“Se potessi”, “Io sono”, “La tua” , “Un giorno”
EuroParlIta	“Dato il”, “Si dovrebbe”,“Penso che”, “In questo”
EuroParlEng	“It is”, “I would”,“You did”, “In this”

Table 7: Prefix sentences used during PPLM generation. We generate three different sentences for each combination of prefix and conditioning label.

References

1. Bianchi, F., Terragni, S., Hovy, D.: Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. arXiv preprint arXiv:2004.03974 (2020)
2. Bianchi, F., Terragni, S., Hovy, D., Nozza, D., Fersini, E.: Cross-lingual contextualized topic models with zero-shot learning. arXiv preprint arXiv:2004.07737 (2020)
3. Blei, David M.; Ng, A.Y.J.M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research pp. 993–1022 (2003)
4. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners (2020)
5. Dathathri, S., Madotto, A., Lan, J., Hung, J., Frank, E., Molino, P., Yosinski, J., Liu, R.: Plug and play language models: a simple approach to controlled text generation. arXiv preprint arXiv:1912.02164 (2019)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. pp. 4171–4186 (2019)
7. Fenu, C.: Sentiment Analysis d'autore: l'epistolario di Italo Svevo. In: Proceedings of 2017 AIUCD 6th Conference on "The educational impacts of DSE". pp. 149–155 (2017)
8. Francia, S., Parisi, L., Paolo, M.: UmBERTo: an Italian Language Model trained with Whole Word Masking. `https://github.com/musixmatchresearch/umberto` (2020)
9. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification (2018)
10. Keskar, N.S., McCann, B., Varshney, L., Xiong, C., Socher, R.: CTRL - A Conditional Transformer Language Model for Controllable Generation. arXiv preprint arXiv:1909.05858 (2019)
11. Koehn, P.: Europarl: A Parallel Corpus for Statistical Machine Translation. In: Conference Proceedings: the tenth Machine Translation Summit. pp. 79–86. AAMT, AAMT, Phuket, Thailand (2005), `http://mt-archive.info/MTS-2005-Koehn.pdf`
12. Li, J., Jia, R., He, H., Liang, P.: Delete, retrieve, generate: a simple approach to sentiment and style transfer. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). pp. 1865–1874. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). https://doi.org/10.18653/v1/N18-1169, `https://www.aclweb.org/anthology/N18-1169`
13. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A robustly optimized BERT pretraining approach (2019)
14. Mattei, L.D., Cafagna, M., Dell'Orletta, F., Nissim, M., Guerini, M.: Geppetto carves italian into a language model (2020)
15. Radford, A.: Improving language understanding by generative pre-training (2018)
16. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners (2019)
17. Sarti, G.: Svevo epistolary corpus analysis. `https://github.com/gsarti/svevo-letters-analysis` (2019)
18. Srivastava, A., Sutton, C.: Autoencoding variational inference for topic models. arXiv preprint arXiv:1703.01488 (2017)
19. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Brew, J.: Huggingface's transformers: State-of-the-art natural language processing. ArXiv **abs/1910.03771** (2019)