

PERSONALIZED DATA-TO-TEXT NEURAL GENERATION

PENS - Stage 2022 - Proposition de méthodologie

19 juin 2023

Ben KABONGO

Stage - M1 DAC - Sorbonne Université

Sommaire

- PENS : Ao et al. 2021
- Résultats du stage 2022
- Reformulating Unsupervised Style Transfer as Paraphrase Generation : Krishna, Wieting et Iyyer 2020
- Proposition de méthodologie

PENS

A Dataset and Generic Framework for Personalized News Headline Generation

- Génération de titres personnalisés d'articles de journaux
- Titre spécifique à l'utilisateur
- En fonction des intérêts de l'utilisateur et du corps de l'article

Dataset

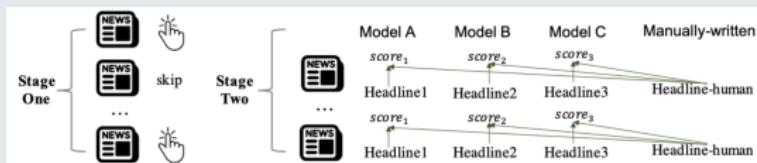
- **Articles** : identifiant, catégorie, sujet, titre, contenu, entités nommées dans le titre et dans le contenu de l'article.
- **Apprentissage et validation** : logs d'impression et l'historique des clics des utilisateurs sur les articles.
- **Test** : 103 locuteurs natifs anglais : annotations des préférences et annotations des titres des articles.

PENS Dataset

Données d'apprentissage et de validation

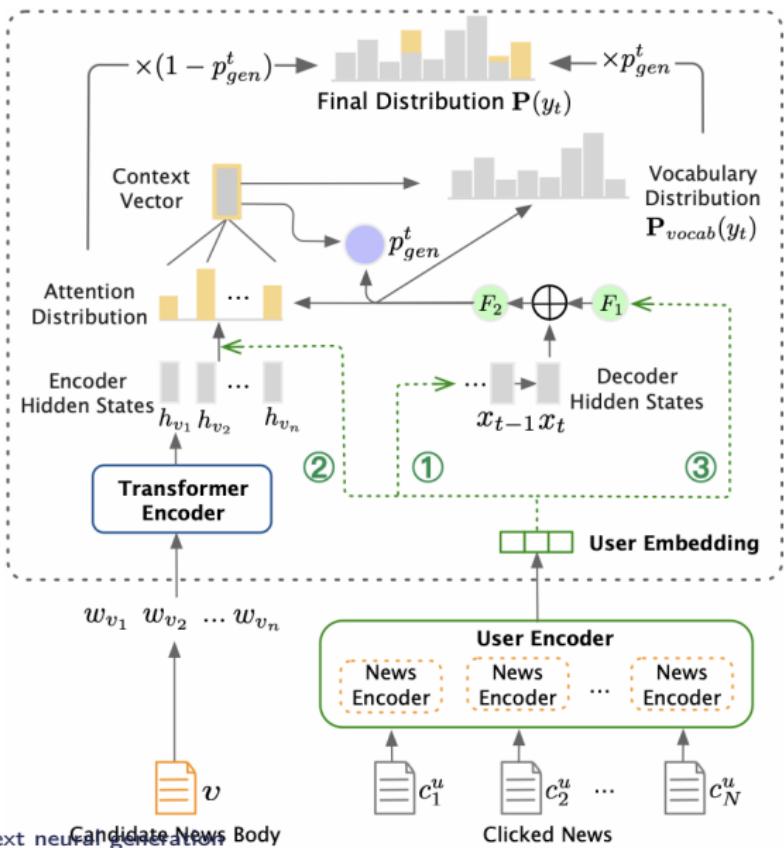
- Logs d'impressions capturées par utilisateur
- Identifiant de l'utilisateur
- L'historique des articles vus par l'utilisateur
- Articles cliqués et non cliqués au moment de l'impression

Données de test



- Annotations d'intérêts sur 1000 articles
- Annotations de préférences des titres de 200 articles

PENS Framework



PENS Framework

Framework

- **Architecture** : transformer : encodeur + décodeur (pointer network)
- **Méthodes d'injection des préférences utilisateurs**
 - 1 Initialiser l'état caché du décodeur du générateur de titres avec l'embedding de l'utilisateur
 - 2 Personnaliser les valeurs d'attention sur les mots dans le corps de l'article par l'embedding de l'utilisateur
 - 3 Perturber le choix entre la génération et la copie grâce à l'embedding de l'utilisateur

Apprentissage

- **Processus de décision markovien** : génération de titres d'articles -> apprentissage par renforcement.
- **Reward** : $\mathbb{E}_{Y_{1:T} \sim G_\theta}[R(Y_{1:T})]$

Résultats

Table 2: The overall performance of compared methods. “R-1, -2, -L” indicate F scores of ROUGE-1, -2, and -L, and “NA” denotes “Not Available”. “IM” means injection methods, c.f. ①, ②, and ③ in Fig. 3 for details.

Methods	Metrics							
	AUC	MRR	NDCG@5	NDCG@10	IM	ROUGE-1	ROUGE-2	ROUGE-L
Pointer-Gen	NA	NA	NA	NA	NA	19.86	7.76	18.83
PG+RL-ROUGE	NA	NA	NA	NA	NA	20.56	8.42	20.03
EBNR	63.97	22.52	26.45	32.81	①	25.13	9.03	20.73
					②	25.49	9.14	20.82
					③	24.62	8.95	20.40
DKN	65.25	24.07	26.97	34.24	①	25.97	9.23	20.92
					②	27.48	10.07	21.81
					③	25.02	8.98	20.34
NPA	64.91	23.65	26.72	33.96	①	25.49	9.14	20.82
					②	26.11	9.58	21.40
					③	26.35	9.71	21.82
NRMS	64.27	23.28	26.60	33.58	①	24.92	9.01	20.75
					②	26.15	9.37	21.03
					③	25.41	9.12	20.91
LSTUR	62.49	22.69	24.71	32.28	①	23.71	8.73	21.13
					②	24.10	8.82	20.73
					③	23.11	8.42	20.38
NAML	66.18	25.51	27.56	35.17	①	27.49	10.14	21.62
					②	28.01	10.72	22.24
					③	27.25	10.01	21.40

Conclusions

Exemple

Table 3: A case study on personalized headline generation for two different users by personalized (NAML+HG) and non-personalized (Pointer-Gen). Underlined words and colored words represent the correlated words in clicked news and the generated headline, respectively.

Case 1. Original Headline:	Venezuelans rush to Peru before new requirements take effect
Pointer-Gen:	Venezuelans rush to Peru
NAML+HG for user A:	Peru has stricter entry requirements for escaping Venezuelans on that influx.
Clicked News of user A:	<ol style="list-style-type: none"> 1. Peru and Venezuela fans react after match ends in a draw 2. Uruguay v. Peru, Copa America and Gold Cup, Game threads and how to watch
NAML+HG for user B:	Stricter entry requirements on Venezuelan migrants and refugees .
Clicked News of user B:	<ol style="list-style-type: none"> 1. Countries Accepting The Most Refugees (And Where They're Coming From) 2. Venezuelan mothers, children in tow, rush to migrate

Conclusions : meilleures méthodes

- NAML
- La méthode d'injection (2)

Stage 2022

Dataset WikiRoto

Klaus



Release poster

Directed by Sergio Pablos

Screenplay by Sergio Pablos
Jim Mahoney
Zach Lewis

Story by Sergio Pablos

Produced by Jinko Gotoh
Sergio Pablos
Marisa Roman
Matt Teevan
Mercedes Gamero
Mikel Lejarza
Gustavo Ferrada

Starring Jason Schwartzman
J. K. Simmons
Rashida Jones

Klaus is a 2019 Spanish-American [animated Christmas film](#) written and directed by [Sergio Pablos](#) in his directorial debut,^[2] produced by his company [Sergio Pablos Animation Studios](#) and distributed by [Netflix](#). Co-written by Zach Lewis and Jim Mahoney, and co-directed by Carlos Martinez Lopez, the traditionally animated film stars the voices of [Jason Schwartzman](#), [J. K. Simmons](#), [Rashida Jones](#), [Will Sasso](#), Neda Margrethe Labba, Sergio Pablos, [Norm Macdonald](#) (in his final film role released in his lifetime), and [Joan Cusack](#). Serving as an alternate origin story of [Santa Claus](#) independent from the historical [Saint Nicholas of Myra](#) and using a fictional 19th-century setting, the plot revolves around a postman stationed in an island town to the Far North who befriends a reclusive toymaker (*Klaus*).

CRITIC REVIEWS FOR *KLAUS*

[All Critics \(74\)](#) | [Top Critics \(16\)](#) | [Fresh \(70\)](#) | [Rotten \(4\)](#)

 The wonderful Rashida Jones voices a teacher who could become a love interest for Jesper. And any movie that casts Norm Macdonald as a cynical boat captain who sounds exactly like Norm MacDdnald is all right by me.

December 3, 2019 | Rating: 3/4 | [Full Review...](#)



Richard Roeper
Chicago Sun-Times
★ TOP CRITIC

 It's awkward and weird, and yet all that awkwardness and weirdness give it personality and charm and a freewheeling, nonsensical quality that feels refreshing.

November 25, 2019 | [Full Review...](#)



Bilge Ebiri
New York Magazine/Vulture
★ TOP CRITIC

 Though there are some narrative

 None of the characters in *Klaus*

Dataset WikiRoto

Construction du dataset

- Inspirée du papier PENS
- Informations sur des films et des critiques utilisateurs :
Wikipedia + Rotten Tomatoes movies (Kaggle)
- Data : infobox des pages Wikipedia des films
- Target data-to-text : premier paragraphe de texte de la page Wikipedia du film
- **Rotten Tomatoes movies** : dataset Kaggle avec des avis de films
- 1130017 avis pour 17712 films et 11108 utilisateurs
- 3277 utilisateurs considérés : ceux ayant laissé au moins 30 avis

Dataset WikiRoto

Structure

- **Features** : 1 exemple = 1 input (data) + target (texte) + 1 critique
 - **Input** : infobox du film (sous différentes formes : paire clé-valeur, format xml, etc.)
 - **Target** : Première phrase/paragraphe de la page wikipédia du film
 - **Critique** : Une seule critique (pas d'info sur l'utilisateur)
- **Jeu de données** : Train (13k), test (1884), validation (1000).

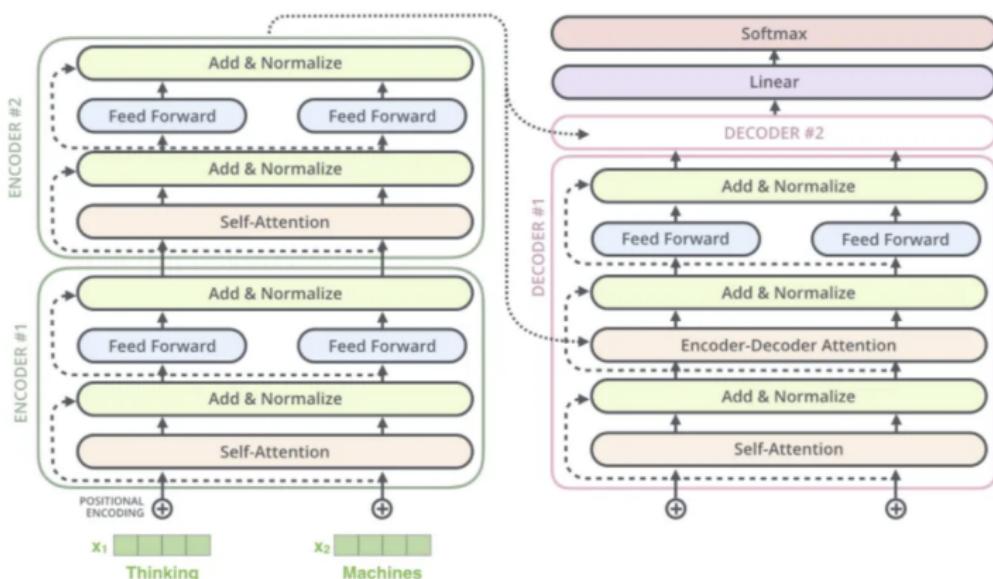
Exemple

- **Input :** <table> <movie_title> Percy Jackson the Olympians : The Lightning Thief </movie_title> <content_rating> PG </content_rating> <genres> Action Adventure, Comedy, Drama, Science Fiction Fantasy </genres> ... </table>
- **Target :** *Percy Jackson the Olympians : The Lightning Thief (also known as Percy Jackson and the Lightning Thief) is a 2010 action fantasy film directed by Chris Columbus*
- **Critique :** *Though it may seem like just another Harry Potter knockoff, Percy Jackson benefits from a strong supporting cast, a speedy plot, and plenty of fun with Greek mythology*

Processus expérimental

- **Data-to-text** : pas de personnalisation
 - Baseline avec des variantes du modèle T5
 - **Données** : input et target du dataset WikiRoto
 - **Loss** : cross-entropy entre le texte généré et la target
 - **Paramètres** : learning rate = 0.001, decay = -0.8, batch size = 16, epochs = 10.
- **Personnalisation** : prompt tuning sur les données de test de PENS
 - Pour chaque utilisateur, sampling de 30 titres dans les préférences, concaténation des titres à insérer dans le prompt.
 - T5-small v1.1, T5-base v1.1 : prompt = le titre de l'article
 - T5 small, T5 base : prompt = corps de l'article + prefix summarize
 - T5 base finetuné sur la génération de titres : prompt = corps de l'article + le préfixe du titre

T5 : Architecture



Métriques

- **BLEU** : Mesure de similarité entre la sortie de la machine et des références d'annotations humaines. Tient compte du nombre de co-occurrences des n-grams ($n \in \{1, 2, 3, 4\}$).
- **PARENT** : Similaire à la métrique BLEU. Mesure à quel point un mot peut être généré étant donné la table des données d'entrée.
- **DIST-N** : Evalue la diversité du texte généré. Nombre de n-grams distincts sur la longueur du texte généré, afin d'éviter de favoriser les longues phrases.

Résultats : Data-to-text

Models	BLEU	PARENT
T5-small	39.60	43.32
T5-base	39.00	42.31
T5-large	39.78	43.76

Table 2: Results on validation set for data-to-text generation

Résultats : Prompt tuning PENS

Models	Prompt length	BLEU	Pairwise DIST-1	Pairwise DIST-2
T5v1.1-small	5	0.18	0.31	0.39
	20	0.23	0.55	0.62
	50	0.49	0.73	0.78
T5v1.1-base	5	0.14	0.53	0.63
	20	0.15	0.42	0.58
	50	0.11	0.56	0.65
T5-small	5	0.48	0.73	0.68
	20	0.36	0.72	0.67
	50	0.28	0.74	0.69
T5-base	5	0.12	0.64	0.69
	20	0.12	0.70	0.80
	50	0.16	0.64	0.69
T5-base finetuned	5	0.37	0.93	0.92
	20	0.52	0.94	0.93
	50	0.48	0.96	0.93
no prompt T5v1.1-small	X	0.17	0.32	0.45
no prompt T5v1.1-base	X	0.14	0.43	0.53
no prompt T5-small	X	0.24	0.68	0.86

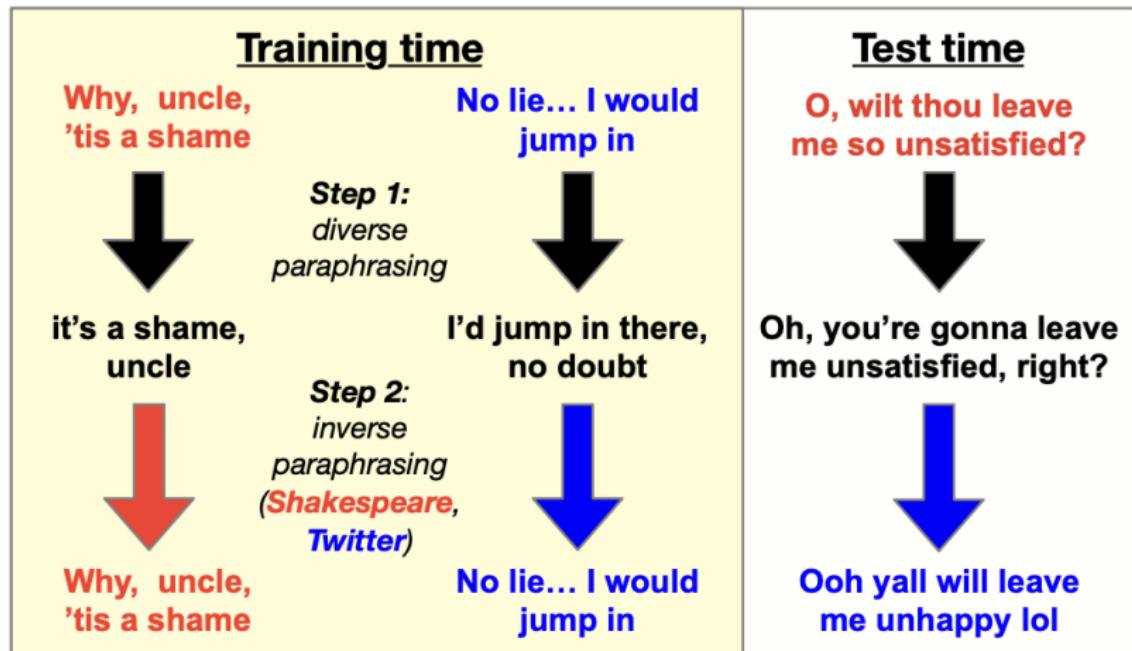
Résultats : Prompt tuning WikiRoto

Prompt type	Prompt length	BLEU	PARENT	Pair-wise DIST-1	Pairwise DIST-2
T5-small v1.1	5	13.21	16.06	0.72	0.77
	20	7.29	10.82	0.68	0.71
	50	4.66	8.76	0.65	0.67
T5-base v1.1	5	17.33	22.49	0.66	0.81
	20	21.59	26.50	0.71	0.85
	50	21.76	25.68	0.70	0.85
T5-small	5	3.91	6.99	0.65	0.68
	20	10.25	13.52	0.67	0.71
	50	14.15	16.28	0.75	0.81
T5-base	5	15.36	19.61	0.69	0.84
	20	18.81	23.74	0.66	0.81
	50	20.61	26.01	0.65	0.79
T5-base finetuned	5	14.07	17.08	0.70	0.85
	20	21.69	26.46	0.70	0.84
	50	16.11	22.40	0.62	0.74
no prompt T5-small	X	32.21	37.30	0.66	0.75
no prompt T5-base	X	30.79	32.94	0.64	0.77

Table 4: Results on WikiRoto validation set for personalized data-to-text generation, mean over 5 users

Reformulating Unsupervised Style Transfer as Paraphrase Generation

Style Transfer via Paraphrasing (STRAP)



Style Transfer via Paraphrasing (STRAP)

- Transfert de style modélisé par génération de paraphrases
- Procédé :
 - **Création de données pseudo-parallèles** pour les différents styles
 - **Entraîner un modèle d'inversion de reformulation** : partir des données de style vers l'entrée initiale
 - **Transfert de style** : par reformulation
- Apprentissage : CDS (Corpus of Diverse Styles) : 15M phrases en anglais avec 11 styles différents + GPT-2
- Framework :
 - **Création des données d'apprentissage** Style i , corpus de phrases X^i : $z = f_{para}(x)$, avec $x \in X^i$. On obtient le corpus pseudo-parallèle (X^i, Z^i)
 - **Paraphrasage inverse** : $x = f_{inv}^i(z)$ avec $z \in Z^i$

Métriques

- **Transfer accuracy (ACC)** : classifieur de style par fine-tuning de RoBERTa-large
- **Semantic similarity (SIM)** : subword embedding-basé sur le SIM modèle de Wieting et al. (2019)
- **Fluency (FL)** : perplexité donnée par l'accuracy avec un classifieur RoBERTa-large entraîné sur le corpus CoLA (Warstadt et al., 2019),
- **Métrique agrégée (J) :**

$$J(ACC, SIM, FL) = \sum_{x \in X} \frac{ACC(x)SIM(x)FL(x)}{|X|}$$

- **Evaluation humaine**

Résultats : Evaluation par métriques

Model	Formality (GYAFC)					Shakespeare				
	ACC	SIM	FL	GM(A,S,F)	$J(A,S,F)$	ACC	SIM	FL	GM(A,S,F)	$J(A,S,F)$
COPY	5.2	80.1	88.4	33.3	4.2	9.6	67.1	79.1	37.1	7.2
NAÏVE	58.9	38.9	89.1	58.9	7.3	49.9	34.9	78.9	51.6	4.1
REF	93.3	100	89.7	94.2	83.8	90.4	100	79.1	89.4	70.5
UNMT (2019)	78.5	49.1	52.5	58.7	20.0	70.5	37.5	49.6	50.8	14.6
DLSM (2020)	78.0	47.7	53.7	58.5	18.6	71.1	43.5	49.4	53.5	16.3
STRAP ($p = 0.0$)	67.7	72.5	90.4	76.3	45.5	71.7	56.4	85.2	70.1	34.7
STRAP ($p = 0.6$)	70.7	69.9	88.5	75.9	44.5	75.7	53.7	82.7	69.5	33.5
STRAP ($p = 0.9$)	76.8	62.9	77.4	72.0	38.3	79.8	47.6	71.7	64.8	27.5

Table 1: Automatic evaluation of our method STRAP (using different p values for nucleus sampling) against prior state-of-the-art methods (UNMT, DLSM), lower bound baselines (COPY, NAÏVE) and reference sentences (REF). STRAP significantly outperforms prior work, especially on our proposed $J(\cdot)$ metric. GM is the geometric mean.

Résultats : Evaluation humaine

Dataset	Model	ACC	SIM	$J(A,S)$	$J(A,S,F)$
Form.	UNMT	77.3	22.7	14.7	7.3
	DLSM	78.0	24.0	15.3	10.0
	$p = 0.0$	71.3	76.0	54.7	41.3
	$p = 0.9$	79.3	56.7	46.0	28.0
Shak.	UNMT	69.3	20.7	10.0	7.3
	DLSM	65.3	37.3	21.3	9.3
	$p = 0.0$	70.7	79.3	56.0	47.3
	$p = 0.9$	74.7	54.0	38.0	24.7

Table 2: Human evaluation of STRAP with greedy decoding ($p = 0.0$) and nucleus sampling ($p = 0.9$) shows large improvements (**4-5x**) on both the Formality (Form.) and Shakespeare (Shak.) datasets. Details on metric calculations are provided in [Appendix A.10](#).

Proposition de méthodologie

Problématique

- **Data-to-text** : données (semi-)structurées (par exemple : table) vers du texte (description des données).
- **Personnalisation** : texte écrit avec le style (les préférences) d'un utilisateur
- **Dataset** : données (semi-)structurées + description textuelle pour chaque type d'utilisateur
- Pas de dataset annoté pour réaliser la tâche

Méthodologie

- **Construction du dataset :**
 - **Data-to-text dataset** : dataset annoté pour les tâches simples de data-to-text (non personnalisé).
 - **Dataset styles utilisateurs** : corpus de styles d'écritures ou dataset de préférences utilisateur
 - **Création du dataset de data-to-text personnalisé** : création des descriptions de chaque donnée suivant le style d'un utilisateur
- **Création du framework de data-to-text personnalisé** : réseau de neurones entraîné sur un dataset de data-to-text personnalisé

Méthodologie

Construction d'un dataset de data-to-text personnalisé

- Entrée : dataset de data-to-text + dataset de styles (préférences) utilisateur
- Sortie : dataset de data-to-text personnalisé
- Reformulation non supervisée par transfert de style
 - Non supervisée car pas de vérité terrain
 - Reformulation non supervisée = **génération de paraphrases**
 - **Procédé** : créer des paraphrases de la vérité terrain de chaque exemple pour les différents styles d'écriture des utilisateurs
 - **Reformulating Unsupervised Style Transfer as Paraphrase Generation** : papier de référence Krishna, Wieting et Iyyer 2020

Méthodologie

Création du framework de data-to-text personnalisé

- **Données** : dataset de data-to-text personnalisé + dataset de styles (préférences) utilisateur
- **Construction d'embeddings utilisateur** : en fonction du style ou des préférences.
- **Evaluation des embeddings utilisateur construits** Terreau, Gourru et Velcin 2020
- **Création d'un réseau de génération de texte personnalisé** : inspiration des méthodes du papier PENS Ao et al. 2021
- Exploration de différentes méthodes et métriques d'évaluation à chaque étape

Références

-  Ao, Xiang et al. (2021). "Pens : A dataset and generic framework for personalized news headline generation". In : *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, p. 82-92.
-  Krishna, Kalpesh, John Wieting et Mohit Iyyer (2020). "Reformulating Unsupervised Style Transfer as Paraphrase Generation". In : *Empirical Methods in Natural Language Processing*.
-  Terreau, Enzo, Antoine Gourru et Julien Velcin (juill. 2020). "Writing Style Author Embedding Evaluation". In : *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Université de Lyon, Lyon 2, ERIC UR3083, p. 84-93. url : https://doi.org/10.26615/978-954-452-056-4_009.