Transformers pour l'Identification de Thèmes et de Mentions dans des Conversations Téléphoniques Bruitées*

Nicolas Andre[1,*,†], Adrien Racamond[1,†] and Mohamed Morchid[1,†]

[1]Avignon Université, Laboratoire Informatique d'Avignon, 84000 Avignon, France

Abstract

Les systèmes basés sur les réseaux de neurones avec système d'attention, tels que les "transformers", ont atteint des performances prometteuses dans différentes tâches liées au traitement automatique du langage naturel dans des environements réels, tel que la compréhension du langage parlé (SLU). Ces modèles neuronaux basés sur les transformers comme BERT, sont très expressifs et efficaces pour la compréhension du langage parlé. Néanmoins, la tâche de détection de thèmes abordés dans des dialogues enregistrés dans des conditions difficiles (téléphone, dialogues enregistrés dans la rue, etc.) s'avère compliquée, car les sujets sont étroitement liés et l'ensemble de dialogues parlés employés pour la phase d'apprentissage est réduit. Cet article propose d'utiliser des modèles de langage basés sur les transformers dans le cadre de la compréhension de dialogues du centre d'appels RATP des données DECODA pour la détection de sujets dans des dialogues parlés bruités. L'article évalue également les performances des transformers pour la détection de sous-thèmes (mentions) portant sur chaque sujet. Les expériences menées montrent que l'utilisation des dépendances entre les mentions et leurs sujets associés améliore à la fois l'identification des thèmes et des mentions. De plus, les expériences soulignent que l'apprentissage des thèmes et des mentions ou sous-thèmes en parallèle permet au système SLU de révéler des dépendances cachées pour un meilleur traitement des appels téléphoniques émanant des clients de la RATP.

Keywords

Transformers, SLU, Classification, Langage

1. Introduction

Les méthodes et algorithmes d'apprentissage automatique ont eu un impact considérable sur un vaste champ de tâches liées à la vie réelle, telles que la reconnaissance vocale [1], l'analyse de la parole [2], les chatbots tels que Alexa d'Amazon [3], la recherche d'informations [4], la détection d'intention[5] [6] [7] ou encore le "slot-filling" [8]. Parmi les techniques d'apprentissage automatique employées, les réseaux de neurones artificiels, et plus précisément les modèles basés sur les système d'attention, ont atteint des performances prometteuses dans diverses tâches de traitement automatique du langage parlé [9] [10]. Les architectures basées sur les transformers comme BERT [11], utilisent des mécanismes d'attention pour capturer les dépendances latentes entre les caractéristiques d'entrée/sortie. Le système BERT a été évalué

CORIA-RJCRI 2024, La Rochelle, France

*Corresponding author.

†These authors contributed equally.

nicolas.andre@univ-avignon.fr (N. Andre); mohamed.morchid@univ-avignon.fr (M. Morchid)

dans différentes tâches liées au traitement du langage, telles que les modèles de langage [12] ou la catégorisation de documents [13], et BERT est rapidement devenu l'état de l'art dans la plupart des tâches liées au traitement automatique du langage naturel (TALN) [14]. Néanmoins, BERT et ses variantes contiennent différentes langues mais utilisent des documents en anglais pour l'apprentissage. Par conséquent, des modèles spécifiques à chaque langue ont été proposés pour surmonter cet inconvénient, comme GottBERT pour l'allemand [15], AlBERTo pour l'italien [16], et FlauBERT [17] ou CamemBERT [18] pour le français. CamemBERT a été évalué sur l'ensemble de données MEDIA [19] pour la classification de texte [20]. Tous ces modèles spécifiques à chaque langue ont été appris avec des ensembles de données plus petits que la version originale de BERT. De plus, les documents utilisés proviennent principalement d'ensembles de données textuelles non bruitées, tels que Wikipedia pour BERT. À notre connaissance, il n'y a pas d'étude sur la robustesse et l'efficacité d'un tel modèle dans des conditions bruyantes avec de petits ensembles de données lors d'une tâche de traitement de la parole. Cet article étudie l'impact des segments bruités dans le signal de parole en raison du contexte d'enregistrement (téléphone, trains, etc.) sur la détection des thèmes et des mentions contenus dans les dialogues parlés issus du corpus DECODA [21] [22] en employant le modèle CamemBERT pour la classification en thèmes et mentions. CamemBERT sera également employé pour la classification multi-thèmes. Un système de reconnaissance automatique de la parole sensible aux erreurs est employé pour maintenir les conditions bruitées d'enregistrement du signal de parole afin d'étudier plus précisément l'impact des segments fortement bruités dans le processus d'apprentissage de CamemBERT. L'article étudie également les dépendances latentes entre les thèmes et les mentions lors des tâches de classification. Le corpus DECODA est composé de conversations entre agent de la RATP et des clients, étiquetées avec un thème correspondant à la préoccupation principale mentionnée par le client ainsi que des mentions liées à ce thème traité dans cette conversation. L'annotation sémantique se compose de 8 thèmes, tels que les problèmes d'itinéraire ou de tarifs, et de 22 mentions telles que "amende" ou "objets trouvés". Il est important de noter que chaque conversation peut contenir plusieurs mentions. Le faible nombre de conversations dans le corpus d'apprentissage de CamemBERT, ainsi que la proximité entre les thèmes, rendent la tâche d'identification des thèmes et des mentions complexe et permet de développer de nouvelles approches pour les applications ne disposant pas de transcriptions peu bruités et de grande quantités de données. Les expériences dans le cadre de DECODA montrent dans un premier temps que les modèles basés sur les transformers, même en français avec un petit nombre de données d'apprentissage, améliorent la précision à la fois lors de la tâche de détection des thèmes ainsi que des mentions contenus dans des dialogues parlés bruités. Les contributions principales de l'article sont les suivantes :

- Evaluer pour la première fois les modèles basés sur les Transformers issus d'une langue spécifique (CamemBERT) sur des documents parlés fortement bruités issus du corpus DECODA lors d'une tâche d'identification de thème.
- Extraire des informations plus précises et pertinentes pour l'agent telles que la mention pour un thème abordé par l'utilisateur lors d'un dialogue DECODA.
- Fusionner le thème et la mention pour améliorer la robustesse des processus d'identification de thème et de mention, avec un gain observé de 3,27 % en termes de précision.

La section 2 donne plus de détails sur l'architecture du modèle et le processus d'apprentissage des tâches d'identification de thème et de mention à partir de documents parlés. Le protocole expérimental et la discussion sur les résultats observés sont détaillés dans les sections 3 et 4 respectivement, et la section 5 conclut cet article et propose un ensemble de perspectives.

2. Approche proposée

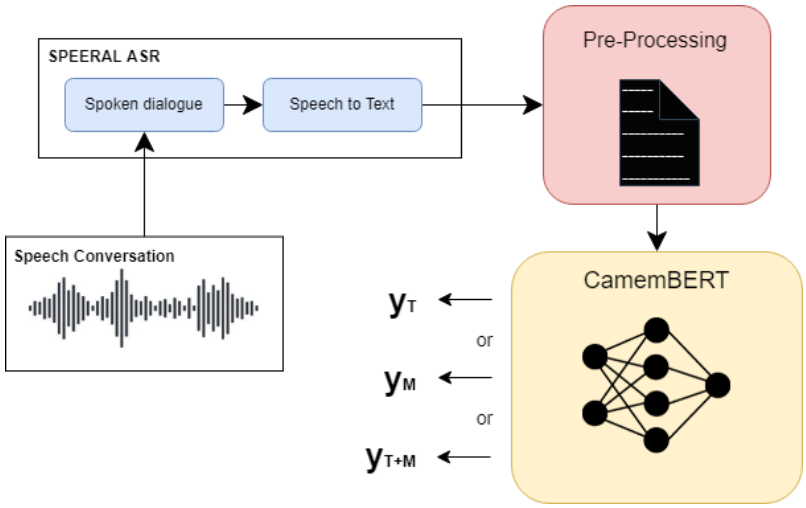

Figure 1: Data processing.

La figure 1 représente l'architecture globale du système pipeline. On emploie, dans un premier temps, un système de reconnaissance automatique de la parole (ASR) pour extraire le contenu du signal parlé dans des transcriptions (documents textuels). Le taux d'erreur global ou "word error rate" (WER) sur l'ensemble de données d'entraînement est de 45,8 % et de 58,0 % sur l'ensemble de test [23]. Plus de détails sur l'ASR employé lors des expériences sont disponibles dans la section 3.1. La transcription est ensuite étiquetée par un humain avec un des 8 thèmes et un ensemble de mentions liées à ce thème principal abordé dans le dialogue. Ces transcriptions sont utilisées par le modèle CamemBERT en entrée et la sortie y du modèle CamemBERT possède différentes longueurs selon la tâche : i) $y_T = 8$ quand la tâche consiste à prédire le thème principal parmi les 8 thèmes composant la sortie ; ii) pour prédire une mention, $y_M = 22$ correspondant aux 22 mentions utilisées pour annoter les dialogues ; iii) $y_{T+M} = 30$ est la concaténation de y_T et y_M et la tâche consiste à prédire le thème ou la mention puisque le processus d'apprentissage considère à la fois les 8 thèmes et les 22 mentions.

2.1. Les modèles de language utilisant les transformers

CamemBERT est un modèle de langue basé sur les transformers. Ces transformers sont des encodeurs-décodeurs basés sur des systèmes d'attention multi-têtes [10]. Étant donné une séquence de requêtes $Q = (Q_1, ..., Q_N)$ et une séquence de clés et de valeurs, $(K, V) =$

$((K_1, V_1)..., (K_N, V_N))$, avec $\boldsymbol{Q} \in \mathbb{R}^{\times \mathbb{T}}$, $\boldsymbol{K}$, $\boldsymbol{V} \in \mathbb{R}^{\times \mathbb{Z}}$, T et D dépendant du modèle, le mécanisme d'attention multi-têtes calcule l'attention entre ces séquences dans plusieurs branches.

$$\text{Scaled Dot-product Attention}(Q, K, V) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}}\right) \times \boldsymbol{V}$$

2.2. RoBERTa

RoBERTa est composé de 12 couches de 768 neurones cachées. Le modèle comprend 3 072 neurones et 12 mécanismes d'attention multi-têtes. Chaque mécanisme d'attention multi-têtes a une taille de 64, et un "dropout" de l'attention de 0,1 est appliquée pendant le processus d'apprentissage avec 24 000 "epochs" et un taux d'apprentissage maximal de 6×10^{-4}. La taille du "batch" est de 8 000, et une décroissance du taux d'apprentissage de 0,01 est appliquée pendant l'apprentissage pour atteindre plus efficacement le point optimal.

2.3. CamemBERT

CamemBERT est un modèle de langage neuronal de langue française à plusieurs couches bidirectionnelles basé sur les transformer [10], similaire à RoBERTa, qui utilise le masquage complet de mots et la tokenisation SentencePiece [24] au lieu de WordPiece [25]. RoBERTa [26] est un modèle de langue anglaise qui est une version modifiée de BERT [27]. CamemBERT a été entraîné sur le corpus français OSCAR [28] composé de 138 Go de texte brut en français, ce qui est significativement plus petit que celui utilisé pour RoBERTa (161 Go de texte anglais). Les paramètres de CamemBERT n'ont pas été ajustés pendant le processus d'apprentissage.

3. Expériences, protocoles et résultats

CamemBERT est utilisé lors d'une tâche de détection de thèmes et de mentions ou sous-thèmes sur des dialogues parlés bruités issus du corpus DECODCA (section 3.1). Cette section explore différents scénarios pour d'abord prédire le thème avec y_T comme sortie du système et la mention avec y_M comme sortie (section 3.2). La deuxième expérience évalue l'impact de la prise en compte des mentions pour prédire le thème ainsi que du thème pour prédire la mention (section 3.3). Enfin, une analyse des résultats est présentée pour mieux comprendre l'impact des mentions étiquetées par les humains sur la tâche d'identification des thèmes dans la section 4.

3.1. Protocole experimental: tâche de SLU

Le corpus DECODA est constitué d'un ensemble de conversations téléphoniques entre les agents et les clients du service client de la RATP. Il comprend 1 242 conversations téléphoniques, représentant 74 heures de signal, transcrites à l'aide du système ASR LIA-Speeral [29] pour préserver le contexte bruité. Chaque conversation a été manuellement transcrite et annotée avec un thème parmi 8 thèmes possibles, correspondant au sujet principal de la conversation, ainsi qu'avec un ensemble de 22 mentions (sous-thèmes) par des humains afin de fournir des informations supplémentaires liées à la conversation. Ces mentions aident l'agent à orienter le client vers le service le plus approprié en fonction de sa demande.

3.2. Prédiction de thèmes et mentions

La première expérience consiste à prédire le thème (y_T) pour un dialogue bruité donné parmi les 8 thèmes. On peut tout d'abord noter, à partir de la figure 2, que la perte observée pendant le processus d'apprentissage atteint relativement rapidement le point optimal (129 epochs sont nécessaires pour l'apprentissage par rapport aux 200 epochs employées comme nombre maximal d'epoch autorisées pour l'apprentissage du modèle). De plus, la précision obtenue est parmi les meilleures observées lors de la tâche d'identification des thèmes sur le cadre DECODA avec 86,1% (voir tableau 2).

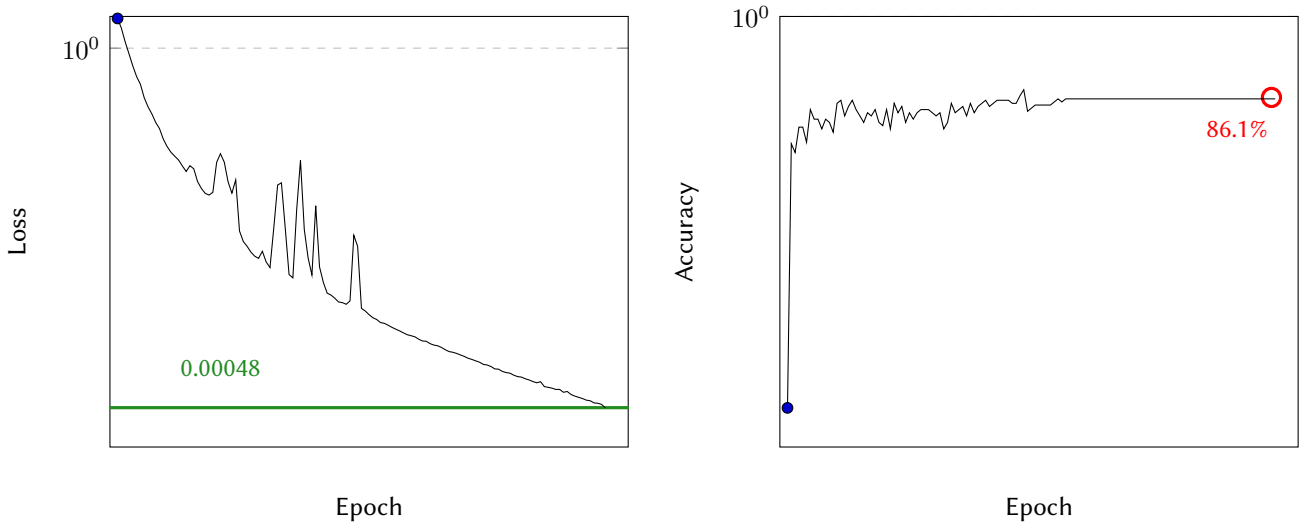

Figure 2: Loss et précision pour les thèmes (l'abscisse est le nombre d'epochs. Sur les 200 epochs, il y a arrêt prématuré à 129 pour éviter le sur-apprentissage).

En effet, le tableau 1 souligne que CamemBERT atteint la meilleure précision sur l'ensemble de données de test, avec un gain de 2,2 points par rapport à l'encodeur-décodeur (AE) et un gain de 1,2 points par rapport à l'encodeur-décodeur profond débruité (DSAE). Le M2H-GAN [30] utilise à la fois des transcriptions provenant de la reconnaissance automatique de la parole des dialogues parlés pour l'apprentissage, ainsi que des documents transcrits par des humains. Par conséquent, les résultats obtenus par M2H-GAN sont spécifiques à la tâche, avec l'utilisation de ressources propres supplémentaires non employées avec CamemBERT.

Table 1
Comparaison de la précision des précedents modèles, sur les données de développement et de test.

Modèles	Pres. Dev	Pres. Test
AE[30]	-	81.0%
DSAE[30]	88.0%	82.0%
CamemBERT$_{BASE}$[18]	**86.1**%	**83.2**%
M2H-GAN[30]	87.0%	85.5%

Les mentions d'un dialogue fournissent des informations plus précises sur l'objet de l'appel

du client et ses besoins. La deuxième expérience extrait de chaque dialogue la mention correspondante y_T parmi les 22 mentions disponibles. En moyenne, un dialogue contient $3,5$ mentions. La précision atteinte lors de cette tâche est de $66,84\%$, ce qui est prometteur puisque toutes les mentions d'un dialogue donné sont proches les unes des autres. De plus, les résultats observés ne sont pas très éloignés de ceux obtenus pour la tâche d'identification des thèmes, car le nombre de mentions est supérieur au nombre de thèmes. Les mentions sont largement influencées par le thème principal abordé dans le dialogue. Par conséquent, les relations latentes entre les thèmes et les mentions doivent être étudiées. La prochaine section évalue l'impact de la prise en compte du thème lors de la prédiction de la mention et l'impact de la mention lors de la prédiction du thème.

3.3. Relations latentes entre thèmes et mentions

La mention fournit une information plus précise qu'un thème et est donc plus utile à l'agent pour orienter l'appel vers le service le plus approprié. Le tableau 2 présente les précisions atteintes par CamemBERT en se basant sur un vecteur de sortie y_{T+M} qui contient à la fois les thèmes et les mentions (taille $= 30$). Cette expérience extrait à la fois la mention et le thème du vecteur de sortie, qui lui-même contient les thèmes et les mentions (y_{T+M}). On peut d'abord noter que la précision obtenue lors de la prédiction de la mention est de $70,127\%$, avec un gain de $3,3$ points par rapport aux $66,84\%$ obtenus précédemment. La précision pour l'identification du thème a baissé à $79,517\%$ avec une perte de $3,5$ points.

Table 2
Précision de la prédiction des mentions avec $y_{T+M} = 30$.

Operation	Pres. Test
Max(y_{T+M}) -> m	**70.127**%
Max(y_{T+M}) -> t	79.517%

Dans l'ensemble, les mentions bénéficient de l'utilisation du thème en raison de la granularité plus large du thème par rapport à la mention. Le thème joue le rôle d'un filtre passe-bas en orientant le système vers une plage plus restreinte de mentions possibles. En revanche, les mentions sont trop spécifiques pour permettre au système basé sur le transformer de trouver le sujet global de la conversation parlée.

4. Analyse des prédictions

La tâche d'identification des thèmes et des mentions des dialogues parlés de DECODA est difficile car les thèmes et les mentions sont souvent proches ou redondants. Par exemple, ITNR (itinéraire) et HORR (horaires d'arrivée) sont souvent sélectionnés à tort. Cela est principalement dû aux mots choisis pour représenter les mentions. Il arrive que les mots associés à la mention apparaissent dans des conversations qui ne sont pas liées au thème principal.

ITNR. Même si l'ensemble de données est petit et que les mentions peuvent être attribuées à tort à un dialogue donné, il s'avère que dans la majorité des cas, nous pouvons identifier

correctement les mentions qui n'auraient pas pu être trouvées en utilisant CamemBERT. Une validation humaine est cependant nécessaire pour les mentions qui pourraient être exprimées avec des mots ou des phrases différents mais similaires. Par exemple, voici une partie d'une conversation qui a été étiquetée comme ITNR (itinéraire) dans le corpus :

.. au depart de la gare antony faire herbe sinon **il faut vous legers vous rendre** dans un et **puis se rendre** a antony chez depart de paris c est le orlybus c est un a oui mais dites deux denfert ..

Par conséquent, l'utilisation des annotations humaines ne suffit pas à extraire, par exemple, la mention "Interaction", car la mention ITNR_Interaction (le client demande une destination, l'agent répond avec la direction à suivre) est correctement prédite par le système.

ETFC. Un autre exemple est la conversation suivante. Elle a été étiquetée manuellement avec le thème ETFC (état du trafic), mais il est très probable que la mention HORR_Horaires (le client demande à l'agent les horaires d'arrivée) soit présente:

bonjour ... un instant s il vous plait oui c est **onze heures** ou la en oui il y a des bus de objets nan c est que le depot de vitry ... c est pour aller a corentin clamart elles sur la ligne **six heures** virement rouge arrete pour aller **vers sept heures dix** les soit dans le bus..

Ci-après, une prédiction correcte pour ETFC est fournie dans une partie d'un dialogue. ETFC_Grève (grève) est prédit et nous pouvons trouver manuellement des mots indiquant un retard sur un train, ce qui pourrait éventuellement être dû à une grève :

blanchard vos oui sur le la il y a des **perturbations rer** c enfin oui bonjour je vous appelle concernant c est des **perturbations sur le rer a son terminus** en si on fait une mais c est le cent vingt

Plus d'attention doit être portée à la prédiction des mentions, car les mentions sont proches en termes de concepts et de mots utilisés pour les décrire.

5. Conclusion

Cet article étudie l'impact du bruit dans des documents parlés lors d'une tâche de détection de thèmes sur le processus d'apprentissage des transformers spécifiques à la langue. Ces modèles sont appris à partir de documents textuels propres tels que Wikipedia pour BERT et, par conséquent, ne conviennent pas aux transcriptions parlées bruitées provenant de signaux parlés enregistrés dans des conditions très bruitées et incontrôlées. Les expériences dans le cadre SLU de DECODA avec le BERT français appelé "CamemBERT" ont montré des performances prometteuses tant pour l'identification des thèmes que des mentions. De plus, l'article souligne que la combinaison des thèmes et des mentions aide à mieux prédire le thème et la mention les plus adaptés. Enfin, l'examen des dialogues parlés incorrectement étiquetés avec une mention révéleque cet étiquetage est souvent lié au thème prédit et à la corrélation entre les thèmes et les mentions.

Les travaux futurs utiliseront des corpus de documents parlés beaucoup plus importants dans des langues différentes de l'anglais pour évaluer l'impact de la taille des ensembles de données pendant le processus d'apprentissage des transformers spécifiques à la langue. De plus, l'impact des langues des ensembles de données utilisés pour l'apprentissage doit également être évalué dans différentes tâches SLU telles que la détection d'intention.

References

[1] M. Malik, M. K. Malik, K. Mehmood, I. Makhdoom, Automatic speech recognition: a survey, Multimedia Tools and Applications 80 (2021) 9411–9457.

[2] P. Vashisht, V. Gupta, Big data analytics techniques: A survey, in: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), IEEE, 2015, pp. 264–269.

[3] V. Kepuska, G. Bohouta, Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home), in: 2018 IEEE 8th annual computing and communication workshop and conference (CCWC), IEEE, 2018, pp. 99–103.

[4] A. Singhal, et al., Modern information retrieval: A brief overview, IEEE Data Eng. Bull. 24 (2001) 35–43.

[5] M. Hardalov, I. Koychev, P. Nakov, Enriched pre-trained transformers for joint slot filling and intent detection, arXiv preprint arXiv:2004.14848 (2020).

[6] C. Xia, C. Zhang, H. Nguyen, J. Zhang, P. Yu, Cg-bert: Conditional text generation with bert for generalized few-shot intent detection, arXiv preprint arXiv:2004.01881 (2020).

[7] H. B. Hashemi, A. Asiaee, R. Kraft, Query intent detection using convolutional neural networks, in: International Conference on Web Search and Data Mining, Workshop on Query Understanding, 2016.

[8] W. A. Abro, G. Qi, M. Aamir, Z. Ali, Joint intent detection and slot filling using weighted finite state transducer and bert, Applied Intelligence (2022) 1–15.

[9] G. Tur, R. De Mori, Spoken language understanding: Systems for extracting semantic information from speech, John Wiley & Sons, 2011.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[11] I. Tenney, D. Das, E. Pavlick, Bert rediscovers the classical nlp pipeline, arXiv preprint arXiv:1905.05950 (2019).

[12] Y. Peng, S. Yan, Z. Lu, Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets, arXiv preprint arXiv:1906.05474 (2019).

[13] H. Shi, C. Wang, Self-supervised document clustering based on bert with data augment, arXiv preprint arXiv:2011.08523 (2020).

[14] F. A. Acheampong, H. Nunoo-Mensah, W. Chen, Transformer models for text-based emotion detection: a review of bert-based approaches, Artificial Intelligence Review 54 (2021) 5789–5829.

[15] R. Scheible, F. Thomczyk, P. Tippmann, V. Jaravine, M. Boeker, Gottbert: a pure german language model, arXiv preprint arXiv:2012.02110 (2020).

[16] M. Polignano, P. Basile, M. De Gemmis, G. Semeraro, V. Basile, Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets, in: 6th Italian Conference on Computational Linguistics, CLiC-it 2019, volume 2481, CEUR, 2019, pp. 1–6.

[17] H. Le, L. Vial, J. Frej, V. Segonne, M. Coavoux, B. Lecouteux, A. Allauzen, B. Crabbé, L. Besacier, D. Schwab, Flaubert: Unsupervised language model pre-training for french,

arXiv preprint arXiv:1912.05372 (2019).
[18] L. Martin, B. Muller, P. J. O. Suárez, Y. Dupont, L. Romary, É. V. de La Clergerie, D. Seddah, B. Sagot, Camembert: a tasty french language model, arXiv preprint arXiv:1911.03894 (2019).
[19] H. Bonneau-Maynard, S. Rosset, C. Ayache, A. Kuhn, D. Mostefa, Semantic annotation of the french media dialog corpus., in: InterSpeech, 2005, pp. 3457–3460.
[20] S. Ghannay, A. Caubrière, S. Mdhaffar, G. Laperrière, B. Jabaian, Y. Estève, Where are we in semantic concept extraction for spoken language understanding?, in: International Conference on Speech and Computer, Springer, 2021, pp. 202–213.
[21] F. Bechet, B. Maza, N. Bigouroux, T. Bazillon, M. El-Beze, R. De Mori, E. Arbillot, Decoda: a call-centre human-human spoken conversation corpus, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), 2012, pp. 1343–1347.
[22] C. Lailler, A. Landeau, F. Béchet, Y. Estève, P. Deléglise, Enhancing the ratp-decoda corpus with linguistic annotations for performing a large range of nlp tasks, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 1047–1050.
[23] M. Morchid, R. Dufour, M. Bouallegue, G. Linares, R. D. Mori, Theme identification in human-human conversations with features from specific speaker type hidden spaces, in: Fifteenth annual conference of the international speech communication association, 2014.
[24] T. Kudo, J. Richardson, Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing, arXiv preprint arXiv:1808.06226 (2018).
[25] M. Schuster, K. Nakajima, Japanese and korean voice search, in: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2012, pp. 5149–5152.
[26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).
[27] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[28] P. J. O. Suárez, B. Sagot, L. Romary, Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures, in: 7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7), Leibniz-Institut für Deutsche Sprache, 2019.
[29] G. Linares, P. Nocéra, D. Massonie, D. Matrouf, The lia speech recognition system: from 10xrt to 1xrt, in: International Conference on Text, Speech and Dialogue, Springer, 2007, pp. 302–308.
[30] T. Parcollet, M. Morchid, X. Bost, G. Linarès, M2h-gan: A gan-based mapping from machine to human transcripts for speech understanding, arXiv preprint arXiv:1905.01957 (2019).