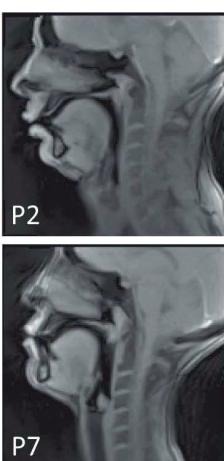
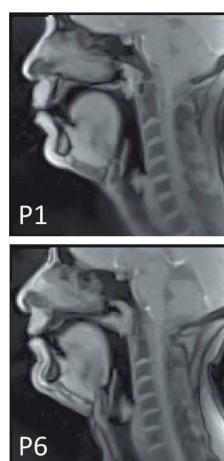
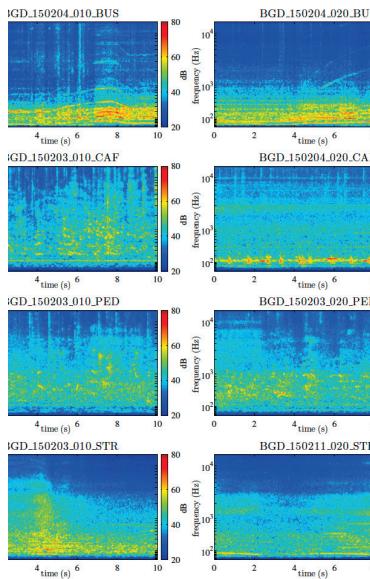
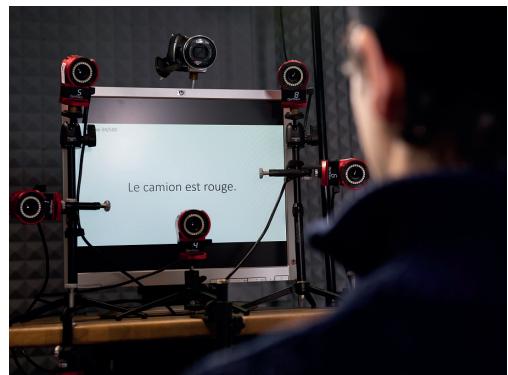


# PORTFOLIO

# DEPARTMENT 4

## Natural Language Processing and Knowledge Discovery



En partenariat avec :



Loria



Inria

UNIVERSITÉ  
DE LORRAINE



# Portfolio Département 4

Nous présentons dans le portfolio du département 4 “Traitement Automatique des Langues et des Connaissances” les 7 documents suivants.

**Rapport détaillé.** L’Université de Lorraine a demandé aux laboratoires de rédiger pour juin 2021 les bilans des laboratoires en vu de l’évaluation HCERES 2016-2020, ceci sans attendre les recommandations de l’HCERES. Nous avons donc rédigé un rapport complet (en anglais), structuré sur la base des anciennes évaluations de l’HCERES. Ce rapport contient une première partie qui présente une vision générale du département et de ses réalisations, suivi d’un rapport détaillé pour chacune des équipes au sein du département. Nous avons naturellement réutilisé pour le DAE une partie substantielle de la première partie, ce qui explique la présence de parties en anglais et en français dans le DAE. Nous présentons ce rapport complet dans le portfolio.

Nous présentons les six autres éléments suivant du portfolio dans la suite de ce document.

- L’article **Predicting the need for a reduced drug dose, at first prescription.** publié dans le journal “Scientific Reports” en 2018)
- L’article **An analysis of environment, microphone and data simulation mismatches in robust speech recognition.** publié dans le journal “Computer Speech and Language” (ISCA Award for the best paper published in Computer Speech and Language (2016-2020))
- L’article **Augmenting Transformers with KNN-Based Composite Memory for Dialog** publié dans le journal “Transactions of the Association for Computational Linguistics”
- La plateforme **Multimodal Speech acquisition**
- La chaire ANR IA xNLG, "**Multi-lingual, Multi-Source Text Generation**"
- les activités **Modélisation des dialogues entre cliniciens et patients souffrant de troubles mentaux**

# Predicting the need for a reduced drug dose, at first prescription

## Prédire la nécessité de réduire le dosage des médicaments, avant la première prescription

<https://hal.inria.fr/hal-01901566>

Équipes K et Orpailleur

### Hétérogénéité de réponses aux médicaments et effets indésirables

Les réponses de notre organisme à un traitement dépendent de nombreux facteurs : notre état général (si nos reins fonctionnent mal, ils élimineront moins bien les médicaments), les traitements que nous suivons déjà (ils peuvent interagir avec les nouveaux traitements), les aliments que nous consommons (le pamplemousse peut par exemple inhiber les enzymes intervenant dans le métabolisme des médicaments), ou notre génome (certains individus ont un déficit partiel – ou parfois total – en une enzyme intervenant dans l'élimination du 5-fluorouracile, un anticancéreux essentiel utilisé pour dans le traitement de nombreux cancers). Pour ces différentes raisons, une dose normale de médicament pour un patient peut s'avérer violemment toxique, voire mortelle, pour d'autres. Pour donner un ordre d'idée, chaque année, aux États-Unis, on estime que 280 000 hospitalisations sont dues aux effets indésirables des médicaments [1]. En France, ces derniers pourraient être responsables de plus de 130 000 cas par an, selon certaines estimations.

L'hétérogénéité des réponses aux médicaments est un phénomène complexe et si les médecins ont l'habitude de considérer le poids, l'âge et l'état général de leurs patients, il leur est difficile de considérer l'ensemble des variables qui pourraient intervenir. Le développement des dossiers de santé numériques (comme les dossiers patients informatisés ou Mon Espace Santé en France) qui conservent les données de traitement des patients, permettent d'imaginer des outils informatiques qui aident les médecins à prescrire de façon plus précise et plus adaptée à chaque patient, au regard à la fois de l'historique propre au patient, mais également de celui des autres patients [3].

### Réutiliser les données de santé

Nous nous sommes intéressés aux médicaments qui interagissent avec les enzymes de la famille des cytochromes P450. Ces enzymes jouent un rôle important dans l'élimination des médicaments, or leur activité varie beaucoup en fonction de la génétique des patients. Elles sont de ce fait à l'origine de réponses très variables d'un patient à l'autre. Nous avons utilisés les dossiers patients électroniques de l'hôpital de l'Université Stanford, en Californie [4]. Celles-ci correspondent aux informations de plus de 1,2 million de patients qui y sont entrés entre 2008 et 2014. Soit un ensemble de près de 50 000 visites s'étant traduites par plus de 20 millions de prescriptions, dont près de 3 millions pour des médicaments ciblant les enzymes P450. Ces données sont anonymisées, puis entreposées à des fins de recherche, chaque patient donnant son accord en amont. En France aussi, le dossier médical partagé gagne chaque jour du terrain : l'Hôpital européen Georges Pompidou a mis en place dès son ouverture, en 2001, un système informatique afin que les dossiers des patients soient électroniques. Au CHRU de Nancy, les dossiers sont également informatisés, et le développement d'un entrepôt de données dédié à leur réutilisation par les chercheurs est en cours de développement. L'accès à ces données à des fins de recherche est conditionné par les autorisation des comités scientifiques et éthiques, les Institutional Review Board aux États-Unis et doivent, en France, respecter le Règlement Général européen sur la Protection des Données (RGPD).

## Développer des modèles prédictifs pour le dosage des médicaments

À partir de ces informations, nous avons, pour 34 médicaments différents, construit des profils de patients [2]. Ces profils regroupent les motifs diagnostiques de l'hospitalisation (insuffisance pulmonaire, greffe de moelle osseuse, transplantation, etc.), les problèmes rencontrés et mentionnés dans les notes cliniques associées (pneumonie, toux sèche, symptôme cutané, etc.) ainsi que les demandes d'analyses en laboratoire. Ces informations ont été croisées avec le devenir de la prescription initiale : a-t-elle nécessité une diminution de dose ? Une augmentation ? Est-elle restée constante ? Ici nous faisons l'hypothèse qu'un changement de dose est un marqueur indiquant la survenue d'un effet indésirable ou d'un manque d'efficacité.

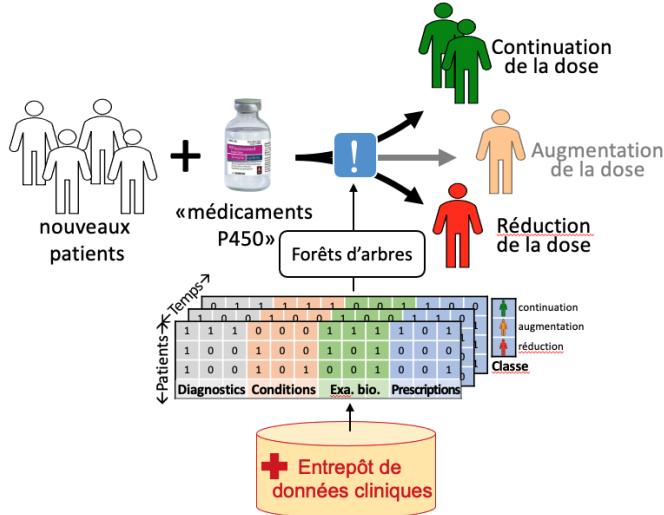


Figure 1: Vue générale du développement et de l'utilisation de nos modèles prédictifs de changement de dose des médicaments P450. Nos modèles sont des forêts d'arbres aléatoires entraînés à partir de données de dossiers patients électroniques. Leur objectif est d'évaluer la nécessité pour un nouveau patient d'avoir une dose réduite en première intention.

Ces données ont été utilisées pour entraîner deux modèles, l'un destiné à prédire les réductions de dose, l'autre les augmentations. Ces modèles permettent alors de statuer sur le fait qu'un nouveau patient va nécessiter dans les semaines à venir d'une réduction, augmentation, ou continuation de la dose (voir Figure 1). Nous avons pour cela employé une méthode couramment utilisée, la méthode Random Forest, ou forêt d'arbres aléatoires. Cet algorithme construit de multiples arbres de décision, chacun entraîné sur un sous-ensemble des données choisi aléatoirement. La décision proposée par chacun des arbres de la forêt est considérée comme un "vote" comptabilisé pour prendre la décision finale.

Nous avons évalué si cette approche pouvait prédire la sensibilité des patients aux médicaments. C'est le cas : elle s'est révélée efficace pour prédire les réductions de doses dans le cas de 23 médicaments sur les 34 étudiés. En effet pour ces 23 médicaments l'AUC ROC est supérieure à 0.7. Elle atteint même 0.9 pour trois médicaments. Parmi ces bons résultats, nous avons isolé les médicaments (14/23) pour lesquels une réduction de la dose n'est motivée (selon les protocoles thérapeutiques) que par une réponse excessive. En revanche notre approche s'est avérée inefficace pour les augmentations de doses ( $AUC ROC < 0.7$  pour tous les médicaments). Afin de faciliter l'interprétation des résultats dans l'optique d'obtenir une interprétation clinique des sensibilités des patients à ces médicaments, nous avons développé une interface Web (<https://snowball.loria.fr/p450/>). Celui-ci permet, sur la bases de mesures comme le risque relatif, le contenu en information et un test statistique, de mettre en avant les attributs les plus importants pour prédire les changements de dose. Cette interface a été utilisé par les médecins de l'équipe de vérifier et proposer une interprétation clinique à nos résultats.

## Intégrer les modèles prédictifs à la pratique clinique

Traditionnellement, les médecins prescrivent les doses de médicaments en suivant des protocoles cliniques : ils débutent un traitement avec une dose de départ, précisée par le protocole, puis évaluent chaque jour ou chaque semaine l'effet du médicament chez le patient afin d'ajuster la dose en fonction d'éventuels effets indésirables. En suggérant l'intérêt d'une dose plus faible, sur la base du profil du patient au moment de la première prescription, notre algorithme peut constituer une aide précieuse pour les médecins. Il les aiderait à prévoir et à prévenir les effets indésirables, réduisant le temps nécessaire pour définir la dose optimale à administrer. Notre approche fonctionne particulièrement bien avec des médicaments utilisés en cancérologie, avec des immunodépresseurs utilisés après une greffe par exemple, ou encore avec des anticoagulants pour lesquels le temps pour atteindre la dose optimale peut-être long, tant la variabilité de réactions des patients est grande. Reste désormais à adapter notre approche aux dossiers électroniques d'hôpitaux français.

## References

- [1] Florence T Bourgeois, Michael W Shannon, Clarissa Valim, and Kenneth D Mandl. Adverse drug events in the outpatient setting: an 11-year national analysis. *Pharmacoepidemiology and drug safety*, 19(9):901–910, 2010.
- [2] Adrien Coulet, Nigam H. Shah, Maxime Wack, Mohammad Chawki, Nicolas Jay, and Michel Dumontier. Predicting the need for a reduced drug dose, at first prescription. *Scientific Reports*, 8(1):1–11, 2018.
- [3] Saurabh Gombar, Alison Callahan, Robert Califf, Robert Harrington, and Nigam H Shah. It is time to learn from patients like mine. *NPJ digital medicine*, 2(1):1–3, 2019.
- [4] Henry J Lowe, Todd A Ferris, Penni M Hernandez, and Susan C Weber. Stride—an integrated standards-based translational research informatics platform. In *AMIA Annual Symposium Proceedings*, volume 2009, page 391. American Medical Informatics Association, 2009.

# An analysis of environment, microphone and data simulation mismatches in robust speech recognition

<https://hal.inria.fr/hal-01399180>

Équipe Multipseech

Today's automatic speech recognition (ASR) systems work well in close-talk scenarios but their performance degrades in hands-free scenarios, especially in noisy environments such as home. To promote solutions to this problem, we have been organizing the series of CHiME Speech Separation and Recognition Challenges since 2011 [2]. Thanks to many academic and industrial research teams, very substantial progress has been made over the years. For the CHiME-3 Challenge in 2015 [1], we provided participants with state-of-the-art system that achieved a word error rate of 33%. A few months later, the best submitted system [5] had reduced it to 6%.

State-of-the-art systems for speech enhancement, separation and ASR are based on deep neural networks trained on tens up to thousands of hours of data. The performance degradation observed in hands-free conditions can therefore be attributed to the mismatched acoustic conditions in the training and test data. Virtually all studies in the literature consider training conditions which match (or cover) the test conditions. In the real world however, the test data often differs from the training data in many ways, including the room's reverberation time, the speaker-to-microphone distance, the intensity of noise, or the noise characteristics (see Fig. 1). In a multichannel setting, the number of microphones, their spatial positions and their frequency response also matter (see Fig. 2).

Besides these mismatches, the mismatch between real and simulated data is also of timely interest. In the era of deep learning, there is an incentive for augmenting the available real training data by perturbing these data or simulating additional training data with similar acoustic characteristics. Simulation might also allow for rough assessment of a given technique in a new environment before real data collected in that environment become available. Suspicion about simulated data used to be common in the speech processing community, due for instance to the misleadingly high performance of certain methods on simulated data compared to real data [4]. Nowadays, this belongs to the past: most modern enhancement and ASR techniques do benefit from data augmentation and simulation [3].

Our study was the first one to systematically review and assess the impact of acoustic mismatches between training and test data on the performance of speech enhancement and ASR systems. We provided a critical analysis of the results published on the CHiME-3 dataset for various signal enhancement, feature extraction, and ASR backend techniques and performed additional experiments in order to separately assess the impact of different noise environments, different numbers and positions of microphones, or simulated vs. real data on speech enhancement and ASR performance. To do so, we exploited a peculiarity of the CHiME-3 dataset: it provides a data simulation tool (entirely designed and implemented by us) which aims to closely reproduce the characteristics of real training data and to generate twinned real and simulated data pairs for development and testing (see Fig. 3). This makes it possible to evaluate the impact of simulated data independently of other changes in the acoustic conditions.

In a nutshell, we showed that, with the exception of minimum variance distortionless response (MVDR) beamforming, most algorithms perform consistently on real and simulated data and can benefit from training on simulated data. We also found that training on different noise environments and different microphones barely affects the ASR performance, especially when several environments are present in the training data: only the number of microphones has a significant impact. Our study received the International Speech Communication Association (ISCA)'s Award for the Best Review Paper published in *Computer Speech and Language* (2016–2020) and it had a significant impact on the community's efforts. As a matter of fact, one year latter, the best word error rate achieved on the CHiME-3 dataset had further dropped to 2% only, and the whole community moved to yet more challenging conditions.

## References

- [1] Jon Barker, Ricard Marxer, Emmanuel Vincent, and Shinji Watanabe. The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines. In *2015 IEEE Workshop on Automatic*

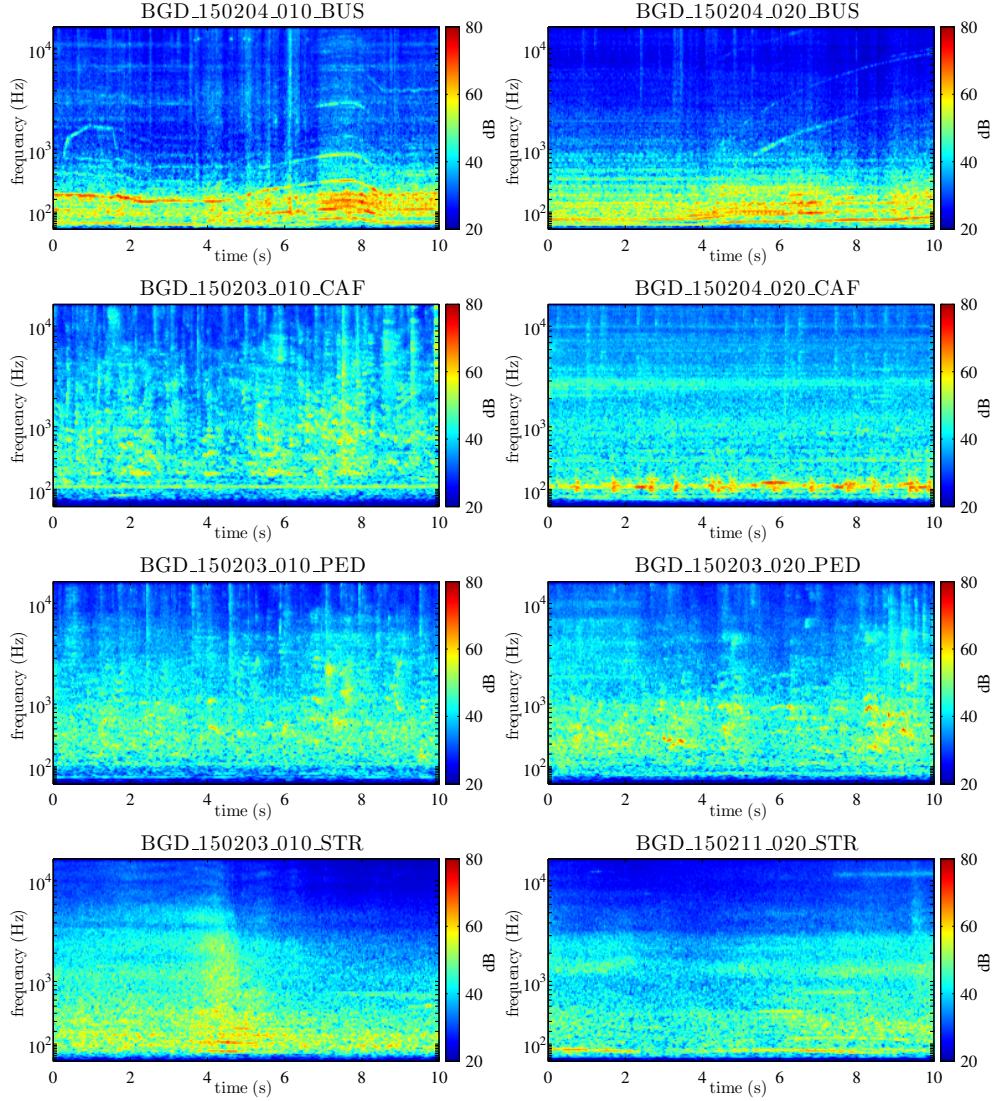


Figure 1: Example spectrograms for two different noise recordings in each environment: bus (BUS), café (CAF), pedestrian area (PED), and street (STR). Many differences can be seen. For instance, BUS noise evolves slowly over time and concentrates below 400 Hz, while PED noise is more nonstationary and wideband. Also, the second instance of CAF noise differs significantly from the first one.

*Speech Recognition and Understanding (ASRU)*, pages 504–511, 2015.

- [2] Jon Barker, Emmanuel Vincent, Ning Ma, Heidi Christensen, and Phil Green. The PASCAL CHiME speech separation and recognition challenge. *Computer Speech & Language*, 27(3):621–633, 2013. Special Issue on Speech Separation and Recognition in Multisource Environments.
- [3] Alessio Brutti and Marco Matassoni. On the relationship between Early-to-Late Ratio of Room Impulse Responses and ASR performance in reverberant environments. *Speech Communication*, 76:170–185, 2016.
- [4] Kenichi Kumatani, Takayuki Arakawa, Kazumasa Yamamoto, John McDonough, Bhiksha Raj, Rita Singh, and Ivan Tashev. Microphone array processing for distant speech recognition: Towards real-world deployment. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–10, 2012.

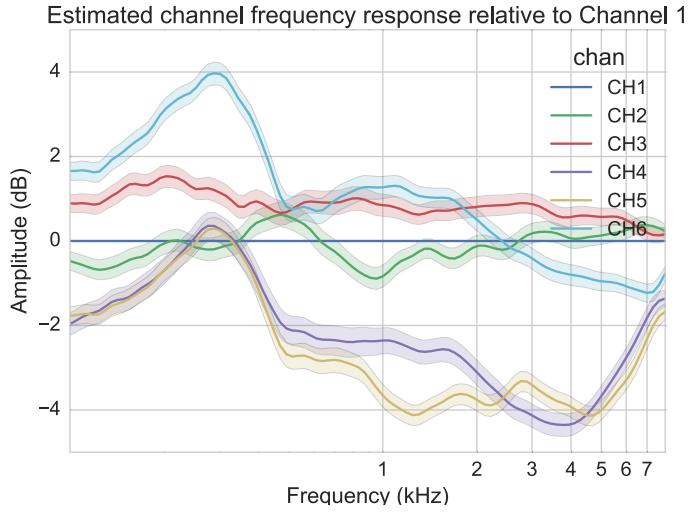


Figure 2: Microphone frequency response relative to channel 1 estimated on 1 min recordings. Solid lines correspond to the mean over the 8 h of continuous noise recordings and colored areas to plus or minus one standard deviation. The data was recorded by a tablet equipped with an array of six microphones, with microphone 2 facing backward on the top center and all other microphones facing forward.

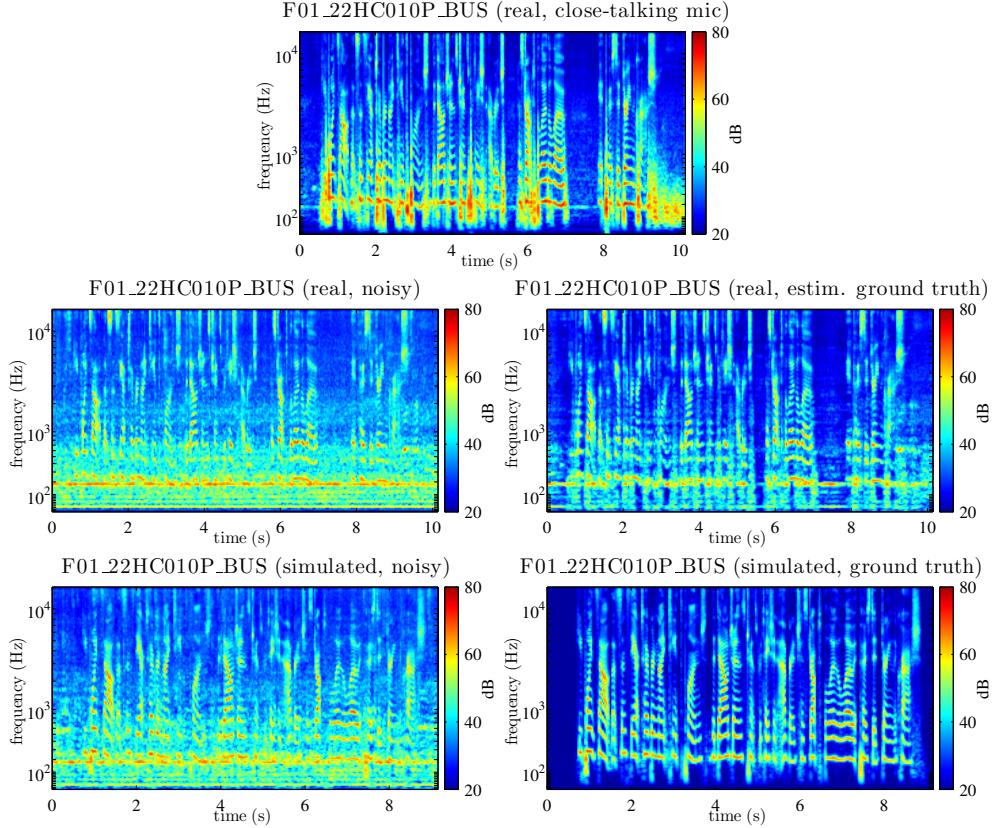


Figure 3: Example real and simulated data and corresponding ground truths.

moto, Chengzhu Yu, Wojciech J. Fabian, Miquel Espi, Takuya Higuchi, Shoko Araki, and Tomohiro Nakatani. The NTT CHiME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 436–443, 2015.

# Augmenting Transformers with KNN-Based Composite Memory for Dialog

<https://hal.inria.fr/hal-02999678>

Équipe Synalp

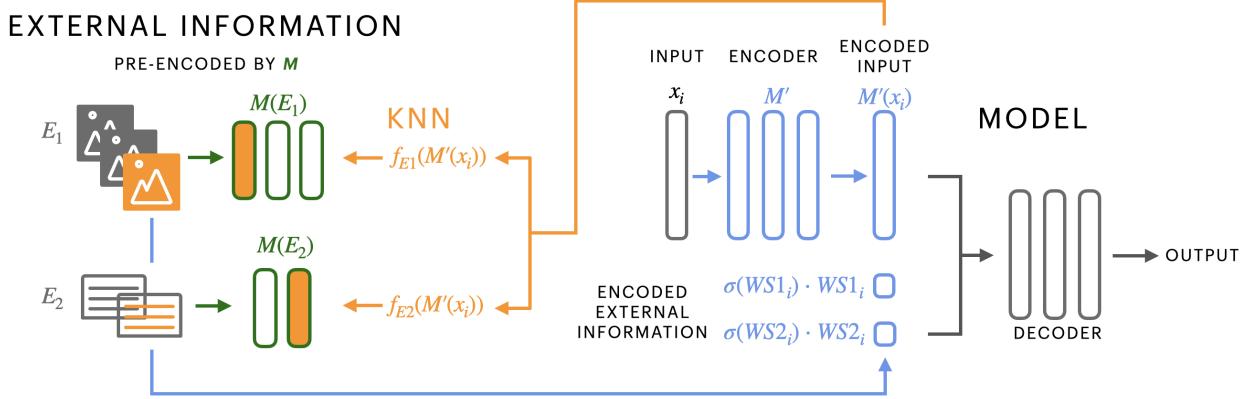


Figure 1: Retrieval-Based NLG. **KIF** modules fetch relevant information from multi-modal external knowledge. External knowledge sources  $E_1$  and  $E_2$  are pre-encoded by encoder  $M$  (green). In the model, input  $x_i$  is encoded by encoder  $M'$  (blue) to produce  $M'(x_i)$ . KIF modules (orange) operate on  $M'(x_i)$  and identify the nearest neighbors encoded in  $M(E_1)$  and  $M(E_2)$  using KNN. Identified relevant elements from  $E_1$  and  $E_2$  are re-encoded by  $M'$  in a gating mechanism with a weighted sum (represented by  $\sigma(\text{WS}1_i) \cdot \text{WS}1_i$ , where WS stands for weighted sum), then concatenated to  $M'(x_i)$ .

NLG tasks, such as human-machine dialog response generation, question-answering, multi-document summarisation or generation of Wikipedia pages, are often dependent on external information. This information can take multimodal forms, including structured knowledge bases, free text or images, and often comes in overwhelmingly large quantities. A pressing challenge is to create models that can identify which specific elements of multiple information sources are relevant in a particular context, and incorporate them into standard architectures on each generation task.

In [3], we focus on human-machine dialog and how to efficiently retrieve relevant external knowledge. We experiment on two dialog tasks, Wizard of Wikipedia [2] and Engaging ImageChat [11]. In both tasks, the generation model must leverage information external to the dialog history alone — in Wizard of Wikipedia, the chat requires access to knowledgeable facts and in Engaging ImageChat, discussion about a specific image. As models must process multiple inputs and ground responses in the knowledgeable facts or images, these tasks challenge existing generation approaches.

To identify relevant pieces of information, previous work has employed architectures that attend over the available knowledge, which scales poorly with large quantities of information [2, 9, 8]. Another line of research has focused on computational scalability for larger external memories to allow efficient access of information. For example, Memory Networks [14, 13, 12] introduce attention mechanisms over large external memories. Neural cache models [4] simplify these to access previous memories with a dot product. Other work has studied how to read and write into these memory architectures [10, 5, 6]. [1] propose a hierarchical memory network rather than a flat one and [10] learn sparse operations to read and write. [7] focus on learning memories of up to one million slots and how to efficiently access the slots using product keys.

In contrast, we propose augmenting generative Transformer neural encoder-decoder networks with KNN-based Information Fetching (KIF) modules. Each KIF module learns a read operation to access fixed external knowledge. Our approach scales well as retrieval benefits from efficient and scalable libraries such as `faiss` and `nmslib`. It allows the retrieved information to be interpreted as each memory slot is an explicit fact that can be read as text, rather than a learned vector such as in [7]. And it flexibly incorporates knowledge from

---

<b>Human:</b> I love watching disney movies! Do you like Disney?	
<b>Model:</b> I love disney, I love watching disney movies and different animations and characters that make them unique	
<b>Fetched Knowledge:</b> <i>disney consumer products and interactive media - the company specializes in producing animated films</i>	
<b>Fetched Training Utterance:</b> <i>i love music, I love going to music shows and dancing and having fun</i>	
<b>Human:</b> What's your favorite disney movie? There are so many good ones.	
<b>Model:</b> I love the incredibles, they are my favorite disney movie.	
<b>Fetched Knowledge:</b> <i>disney announced intentions to develop additional superhero films after the success of the incredibles</i>	
<b>Fetched Training Utterance:</b> <i>i love kiteboarding, it is one of my favorite activities on the water.</i>	
<b>Human:</b> omg, me too! the second one just came out i think. I don't watch a lot of sequels though, I think the last one I saw was toy story.	
<b>Model:</b> That's cool! I remember when toy story was first released in 1995.	
<b>Fetched Knowledge:</b> <i>talk of a second film following toy story began around a month after the film's opening, in december 1995</i>	
<b>Fetched Training Utterance:</b> <i>that's neat! i like so many of her songs! i remember when her first album was released in 2008.</i>	
<b>Human:</b> I've actually never seen toy story before. Kind of embarrassing...	
<b>Model:</b> Ohhh ... I think toy story is a classic? It was first released in 1999.	
<b>Fetched Knowledge:</b> <i>toy story 2 is a 1999 american computer-animated comedy film produced by pixar animation studios for walt disney pictures</i>	
<b>Fetched Training Utterance:</b> <i>Ohhh ... What do people with color blindness do to cope with the effects?</i>	

---

Figure 2: **Conversation between Human and KIF-Augmented Transformer on Wizard of Wikipedia.** The top-scoring Wikipedia knowledge and training utterances fetched by KIF are displayed with model output.

multiple, multi-modal sources. Figure 1 illustrates the architecture of our model when using two separate sources of external knowledge.

We apply our approach to generative dialog modeling, a complex task where relevant information must be identified, retrieved and incorporated to maintain the topic and flow of conversation. As illustrated in Figure 2, we retrieve two types of knowledge to enhance the quality of the dialog engine responses: (i) knowledge about similar dialog contexts and (ii) external knowledge used to ground the conversation into real world information. We demonstrate the effectiveness of our approach by identifying relevant knowledge required for knowledgeable but engaging dialog from Wikipedia, images, and human-written dialog utterances, and show that leveraging this retrieved information improves model performance, measured by automatic and human evaluation.

## References

- [1] Sarath Chandar, Sungjin Ahn, Hugo Larochelle, Pascal Vincent, Gerald Tesauro, and Yoshua Bengio. Hierarchical Memory Networks. *CoRR*, abs/1605.07427, 2016.
- [2] Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. Wizard of Wikipedia: Knowledge-Powered Conversational Agents. In *International Conference on Learning Representations*, 2018.
- [3] Angela Fan, Claire Gardent, Chloé Braud, and Antoine Bordes. Augmenting Transformers with KNN-Based Composite Memory for Dialog. *Transactions of the Association for Computational Linguistics*, 9:82–99, 2021.

- [4] Edouard Grave, Armand Joulin, and Nicolas Usunier. Improving Neural Language Models with a Continuous Cache. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.
- [5] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural Turing Machines. *arXiv preprint arXiv:1410.5401*, 2014.
- [6] Armand Joulin and Tomas Mikolov. Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets. In *Advances in Neural Information Processing Systems*, pages 190–198, 2015.
- [7] Guillaume Lample, Alexandre Sablayrolles, Marc’Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Large Memory Layers with Product Keys. In *Advances in Neural Information Processing Systems*, pages 8548–8559, 2019.
- [8] Rongzhong Lian, Min Xie, Fan Wang, Jinhua Peng, and Hua Wu. Learning to Select Knowledge for Response Generation in Dialog Systems. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5081–5087. AAAI Press, 2019.
- [9] Lianhui Qin, Michel Galley, Chris Brockett, Xiaodong Liu, Xiang Gao, Bill Dolan, Yejin Choi, and Jianfeng Gao. Conversing by Reading: Contentful Neural Conversation with On-demand Machine Reading. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5427–5436, 2019.
- [10] Jack Rae, Jonathan J. Hunt, Ivo Danihelka, Timothy Harley, Andrew W Senior, Gregory Wayne, Alex Graves, and Timothy Lillicrap. Scaling Memory-Augmented Neural Networks with Sparse Reads and Writes. In *Advances in Neural Information Processing Systems*, pages 3621–3629, 2016.
- [11] Kurt Shuster, Samuel Humeau, Hexiang Hu, Antoine Bordes, and Jason Weston. Engaging Image captioning via Personality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12516–12526, 2019.
- [12] Sainbayar Sukhbaatar, Edouard Grave, Guillaume Lample, Herve Jegou, and Armand Joulin. Augmenting Self-attention with Persistent Memory. 2019.
- [13] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-End Memory Networks. In *Advances in neural information processing systems*, pages 2440–2448, 2015.
- [14] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory Networks. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

# Multimodal Speech acquisition

Équipe Multispeech

The multispeech team considers speech as a multimodal signal with different facets: acoustic, facial, articulatory, gestural, etc. Historically, speech has been considered mainly under its acoustic facet. However, the acoustic signal is the consequence of the deformations of the vocal tract under the effect of the movement of the jaw, the lips, the tongue, the soft palate and the larynx to modulate the excitation signal produced by the vocal cords or the aerodynamic turbulence. These deformations are visible on the face (lips, cheeks, jaw) thanks to the coordination of the various orofacial muscles and the deformation of the skin induced by them. Our team has a strong experience in multimodal speech acquisition and the study of articulatory speech production using very advanced techniques. In what follows, we focus on two topics: the multimodal data acquisition platform and the rt-MRI database for French.

**MultiMod: Multimodal Speech acquisition platform.** For several years, we worked on improving acquisition techniques to be able to acquire very good quality data that can capture the finest speech-related movements, and in a reasonable time. We have developed a multimodal acquisition platform called MultiMod. This platform is almost unique in France and in the world by the richness of the systems used: Motion capture with sensors (Vicon, Optitrack) and without sensors (Kinect, RealSense), Articulatory (articulograph 501), Video (camera), audio (microphone). These heterogeneous systems require the development of synchronization and multimodal fusion techniques. We have developed the necessary software tools to make this system operational and to provide good results. We have developed VisArtico [4] which is a multimodal data visualization software acquired by our multimodal platform. This software makes it possible to visualize the positions of real or virtual sensors and to animate them simultaneously with acoustics. It also allows the annotation and editing of the represented data.

This platform has been used to acquire audiovisual data in order to develop our expressive audiovisual speech synthesis system [2] and our multilingual lipsync system [1]. In fact, we are currently developing a realistic talking head where its articulation is finely controlled by a multimodal coarticulation model based on neural networks [1]. It has also been used to study stuttering from an articulatory point of view. This platform is a part of the Creativ' Lab of LORIA.



Figure 1: Multimodal Speech acquisition platform - Illustrating the articulograph, and the Motion capture system.

**Construction of a rt-MRI (real-time Magnetic Resonance Imaging) database for French.** The investigation of the movement of speech articulators has a number of applications including study of speech production, speech recognition, as well as some medical applications: diagnosis and rehabilitation of abnormal speech and swallowing, study of ortho-facial structures implicated in sleep apnoea syndrome. Information on motion can be obtained using different methods including electromagnetic articulography (EMA), X-ray and ultrasound imaging. Nowadays, magnetic resonance imaging (MRI) holds one of the leading positions as a data acquisition method in speech sciences due to its non-invasiveness and absence of long-term health

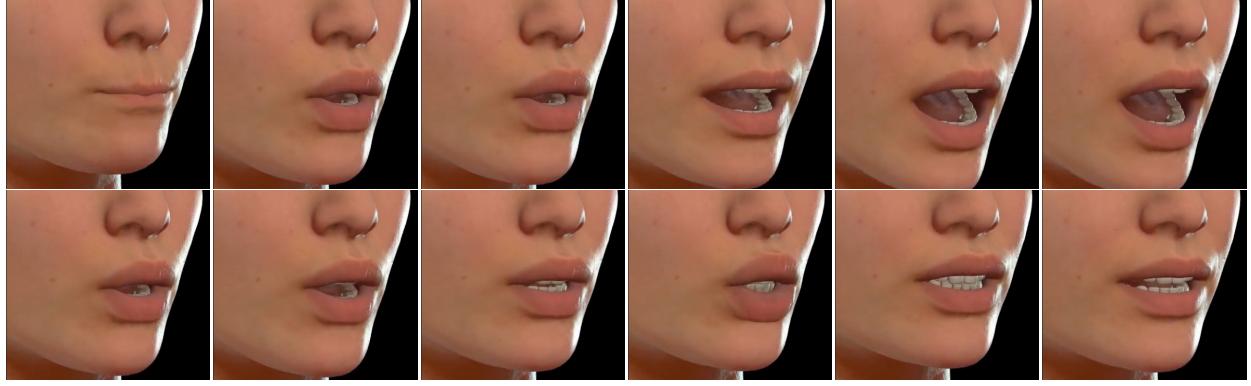


Figure 2: A sequence of phonemes generated by our multimodal coarticulation model [1] to animate a realistic 3D facial model

hazards. Contrarily to other techniques such as ultrasound, which fails to visualize the articulators separated from the sensor by air, or EMA which provides only the sensors' trajectories glued on the upper vocal tract articulators, MRI succeeds to visualize the whole vocal tract.

However, MR imaging of a speaking person is a challenging problem due to the fast motion of articulators. Real-time MRI, in contrast, allows high spatio-temporal resolution. In collaboration with the IADI laboratory, we have created a multimodal MRI database consisting of 2D real-time and 3D static MR images of the vocal tract of 10 French speakers who each pronounced 77 sentences covering all consonantal contexts including the vowels /a,i,u,y/. A real-time MRI technology with temporal resolution of 20 ms was used to acquire vocal tract images of the participants speaking. The sound was recorded simultaneously along with MRI, denoised and temporally aligned with the images. The speech was transcribed to obtain phoneme-wise segmentation of the sound signal. We also acquired static 3D MR images for a wide list of French phonemes. In addition, we included annotations of spontaneous swallowing. This rt-MRI is the first database for French [3].

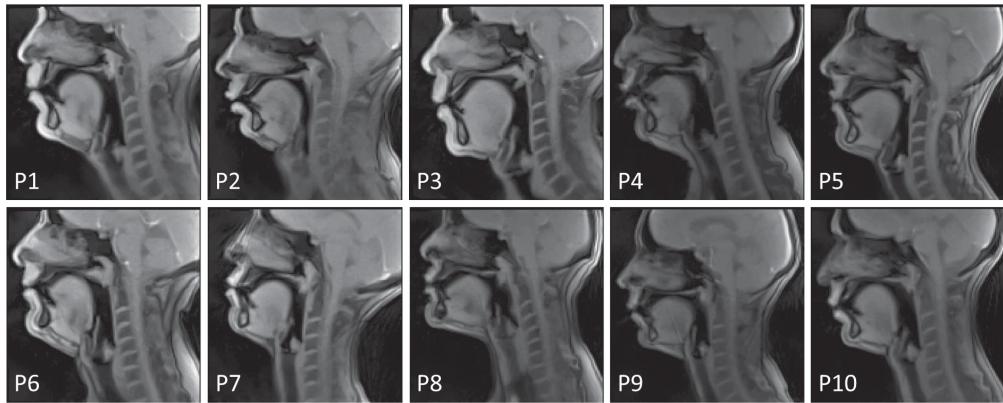


Figure 3: Examples of real-time images of all ten speakers pronouncing /u/.

## References

- [1] Théo Biasutto-Lervat, Sara Dahmani, and Slim Ouni. Modeling Labial Coarticulation with Bidirectional Gated Recurrent Networks and Transfer Learning. In *INTERSPEECH 2019 - 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, September 2019.

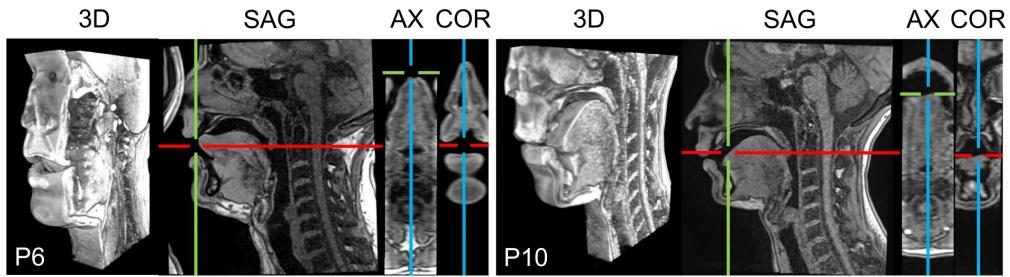


Figure 4: Examples of 3D static images: P6 pronouncing /l/ in context of /a/, and P10 pronouncing /ã/. The 3D volumes were cropped for better visibility.

- [2] Sara Dahmani, Vincent Colotte, Valérian Girard, and Slim Ouni. Learning emotions latent representation with CVAE for Text-Driven Expressive AudioVisual Speech Synthesis. *Neural Networks*, 141:315–329, 2021.
- [3] Karyna Isaieva, Yves Laprie, Justine Leclère, Ioannis K Douros, Jacques Felblinger, and Pierre-André Vuissoz. Multimodal dataset of real-time 2D and static 3D MRI of healthy French speakers. *Scientific Data* , 8(1):258, October 2021.
- [4] Slim Ouni, Loïc Mangeonjean, and Ingmar Steiner. VisArtico: a visualization tool for articulatory data. In *13th Annual Conference of the International Speech Communication Association - InterSpeech 2012*, Portland, OR, United States, September 2012.

# xNLG

## Multi-lingual, Multi-Source Text Generation

funded by the French National Research Agency (ANR-20-CHIA-0003)

Équipe Synalp

While humans can both understand and produce language, most research in Natural Language Processing has focused on Natural Language Understanding (NLU) rather than Natural Language Generation (NLG). One main reason for this is that the input to NLU is well defined (it is text) and easy to access whereas the input to NLG is much more varied. Broadly, it can be (graphical, numerical, symbolic) data, text or meaning representations (MR). By allowing for data verbalisation or summarisation, so-called Data-to-Text (D2T) generation makes such data accessible to lay users. Text-to-text generation is the driving technology behind automatic text summarisation, text simplification, text paraphrasing and text compression. Finally, MR-to-Text generation (generation from meaning representations) is a key part of D2T generation systems. It is used after content selection and content planning to transform the meaning representations produced by these modules into Natural Language sentences. Importantly, it can also be viewed as a sentence level, semantic interlingua between input and output.

Neural methods have introduced a paradigm shift in NLG and considerably broadened NLG research by supporting two key features. First, recurrent neural networks (RNNs) allow for the learning of powerful language models which can be conditioned on arbitrary input (data, text or meaning representations) and are not limited by the Markov assumption. In practice, this proved to support the generation of fluent, natural sounding text. Second, the encoder-decoder architecture provides a natural and unifying framework for all generation tasks independently of the input type (data, text or MR). These two features have led to a veritable explosion of the NLG field with in particular, the creation of several data sets for MR-, Data- and Text-to-text generation.

Building on collaborations with industry (Facebook, Orange, Hugging Face) and academia (U. Darmstadt, Germany; Federal University of Minas Gerais, Brazil; University of Gothenburg, Sweden; Tilburg University, The Netherlands; Universitat Pompeu Fabra, Spain; Paderborn University, Germany), the xNLG Chair focuses on four key challenges for neural NLG: how to generate from large scale data (retrieval-based NLG), how to generate into languages other than English (multilingual NLG), how to verbalise Knowledge-Bases (KB-to-Text generation) and how to simplify documents (Document Simplification).

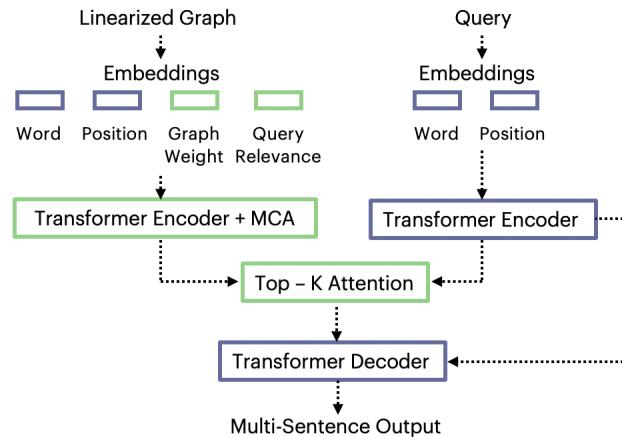


Figure 1: Generating Long Form Text from Large Scale Input

converting the input documents into a graph, we reduce this to 10,000 tokens in average while filtering out inconsistencies and redundancies. We use Memory Compressed Attention mechanism which reduces the

**Retrieval-Based NLG.** NLG tasks, such as dialog, question-answering, multi-document summarisation or generation of Wikipedia pages, are often dependent on external information. This information can take multimodal forms, including structured knowledge bases, free text, and images, and also comes in overwhelmingly large quantities. A pressing challenge is to create models that can identify which specific elements of multiple information sources are relevant in a particular context, and incorporate them into standard architectures on each task. We proposed two approaches for handling large scale input.

In [5], our input are documents retrieved from the web whose average length is 200,000 tokens and the task is either to answer a question from the “Explain Like I am Five” Reddit subforum or to generate a Wikipedia article based on its title. By con-

number of keys and values used in attention computation to encode these 10K tokens in a Transformer network and we leverage the input query (a question for question answering and a Wikipedia article title for summarisation) to learn a sub-selection operation on the input using hierarchical top-k attention. We show that for our two tasks, long-form question answering and multi-document summarization, using graph representations as input achieves better performance than using text fragments.

In [6], we focus on human-machine dialog and how to efficiently retrieve external knowledge that is relevant to the dialog. Some work has focused on learning architectures with large memories capable of storing this knowledge. Instead, we proposed augmenting generative Transformer neural networks with KNN-based Information Fetching (KIF) modules. Each KIF module learns a read operation to access fixed external knowledge. We apply these modules to generative dialog modeling, a challenging task where information must be flexibly retrieved and incorporated to maintain the topic and flow of conversation. We consider two scenarios and for each scenario, retrieve two types of knowledge: (i) knowledge about similar dialog contexts and (ii) external knowledge used to ground the conversation into real world information. We demonstrate the effectiveness of our approach by identifying relevant knowledge required for knowledgeable but engaging dialog from Wikipedia, images, and human-written dialog utterances, and show that leveraging this retrieved information improves model performance, measured by automatic and human evaluation.

**Multilingual NLG.** Generating text from structured data is challenging as it requires bridging the gap between a semantically underspecified structure and a natural language sequence. Multilingual generation brings in an additional challenge: that of generating into languages with varied word order and morphological properties. In [4], we focus on Abstract Meaning Representations (AMRs) as structured input, where previous research has overwhelmingly focused on generating only into English. We leverage advances in cross-lingual embeddings, pretraining, and multilingual models to create multilingual AMR-to-text models that generate in twenty one different languages (Cf. Fig. 2). For eighteen languages, based on automatic metrics, our multilingual models surpass baselines that generate into a single language. We analyse the ability of our multilingual models to accurately capture morphology and word order using human evaluation, and find that native speakers judge our generations to be fluent.

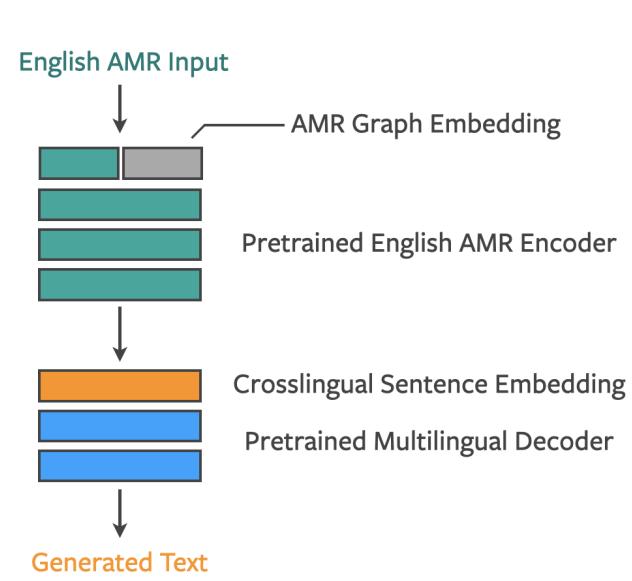


Figure 2: From Abstract Meaning Representations to text in 18 Languages

currently organising a third challenge in collaboration with DCU Dublin and Malta University, focusing on low resource languages (breton, maltese, welsh and irish) i.e., languages for which the generation of high

**Verbalising Knowledge Bases.** While the web makes available large quantities of texts which can be used for summarisation, question answering or dialog modeling, it also provides large knowledge-bases whose access can be facilitated by NLG. In 2017, the SYNALP team organised the first shared task dedicated to the comparison and evaluation of KB verbalisers i.e., NLG models which can convert KB fragments into the corresponding text. xNLG further explores this task, focusing on how to generate from KB data into multiple, in particular, low resource languages. In 2020, we organised a second instance of WebNLG in collaboration with researchers from Germany, Sweden, Brazil, Spain and The Netherlands, this time integrating two languages (English and Russian) and both directions (from text to KB and from KB to Text). This second shared task attracted the participation of 15 teams from both industry (Google, Orange, Facebook, Amazon, Huawei, Sber Bank) and academia (Montreal, Prague, Barcelone, Saigon University). The results showed large improvements with respect to 2017 both for NLU and for NLG [2, 7, 1]. We are

quality text remains an open challenge. xNLG also explored multilingual generation from other forms of structured input such as dependency trees [10, 9, 11] and monolingual generation from knowledge graphs [8].

**Simplifying Text.** Another key challenge we address in xNLG is Document level simplification as most existing work focuses on sentence simplification. A first step in that direction was to provide training data and a model for discourse level splitting i.e., the splitting of a complex sentence into two simpler sentences related by a discourse relation [3]. We are currently working on integrating on a simplification model which recursively simplifies a document using rephrasing, deletion, syntax- and discourse-based sentence splitting.

## References

- [1] Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, and Anastasia Shimorina. The 2020 bilingual, bi-directional WebNLG+ shared task: Overview and evaluation results (WebNLG+ 2020). In *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, pages 55–76, Dublin, Ireland (Virtual), 12 2020. Association for Computational Linguistics.
- [2] Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, and Anastasia Shimorina, editors. *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, Dublin, Ireland (Virtual), 12 2020. Association for Computational Linguistics.
- [3] Liam Cripwell, Joël Legrand, and Claire Gardent. Discourse-based sentence splitting. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 261–273, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [4] Angela Fan and Claire Gardent. Multilingual AMR-to-text generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2889–2901, Online, November 2020. Association for Computational Linguistics.
- [5] Angela Fan, Claire Gardent, Chloé Braud, and Antoine Bordes. Using local knowledge graph construction to scale Seq2Seq models to multi-document inputs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4186–4196, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [6] Angela Fan, Claire Gardent, Chloé Braud, and Antoine Bordes. Augmenting transformers with KNN-based composite memory for dialog. *Transactions of the Association for Computational Linguistics*, 9:82–99, 2021.
- [7] Diego Moussallem, Paramjot Kaur, Thiago Ferreira, Chris van der Lee, Anastasia Shimorina, Felix Conrads, Michael Röder, René Speck, Claire Gardent, Simon Mille, Nikolai Ilinykh, and Axel-Cyrille Ngonga Ngomo. A general benchmarking framework for text generation. In *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, pages 27–33, Dublin, Ireland (Virtual), 12 2020. Association for Computational Linguistics.
- [8] Leonardo F. R. Ribeiro, Yue Zhang, Claire Gardent, and Iryna Gurevych. Modeling global and local node contexts for text generation from knowledge graphs. *Transactions of the Association for Computational Linguistics*, 8:589–604, 2020.
- [9] Anastasia Shimorina and Claire Gardent. LORIA / lorraine university at multilingual surface realisation 2019. In *Proceedings of the 2nd Workshop on Multilingual Surface Realisation (MSR 2019)*, pages 88–93, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [10] Anastasia Shimorina and Claire Gardent. Surface realisation using full delexicalisation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International*

*Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3086–3096, Hong Kong, China, November 2019. Association for Computational Linguistics.

- [11] Anastasia Shimorina, Elena Khasanova, and Claire Gardent. Creating a corpus for Russian data-to-text generation using neural machine translation and post-editing. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 44–49, Florence, Italy, August 2019. Association for Computational Linguistics.

# Modélisation des dialogues entre cliniciens et patients souffrant de troubles mentaux

projet MePheSTO — Inria Sémagramme (Nancy), Star (Nice) et DFKI

action Exploratoire Inria « ODiM »

**Équipe Sémagramme**

Deux grands défis persistent dans la psychiatrie moderne : l'absence de marqueurs quantitatifs clairs de la maladie et le manque de corpus suffisamment importants combinant des mesures multimodales et des évaluations longitudinales [2]. Aujourd'hui, les états cliniques sont mesurés à l'aide d'échelles basées sur des questions liées à des domaines symptomatiques spécifiques qui peuvent être sujettes à de nombreux biais d'interprétation [3]. L'hétérogénéité des symptômes et la variabilité entre les patients constituent un défi qui peut conduire à manquer des diagnostics ou à prescrire des traitements incorrects [4]. En l'absence de marqueurs quantitatifs et objectifs de la gravité de la maladie, il est également difficile de créer des méthodes efficaces et efficientes pour définir des stratégies de traitement médicamenteux adaptés. Par exemple, dans le cas des troubles bipolaires, les stabilisateurs d'humeur (lithium, anticonvulsivants, antipsychotiques) constituent le traitement principal pour la prévention des rechutes à long terme, mais leurs effets secondaires peuvent entraîner une réduction significative de la qualité de vie des patients. L'identification précoce de la rechute, ainsi que l'ajustement de la médication et de la psychothérapie apparaissent comme des solutions plus efficaces.

Pour développer et tester des marqueurs quantitatifs et objectifs de la maladie mentale, une solution est de passer des marqueurs et du phénotypage traditionnels aux marqueurs numériques et au phénotypage numérique. Le « phénotypage numérique » fait référence à la quantification instantanée du comportement humain dans la vie quotidienne à l'aide de données provenant d'appareils numériques [6]. Le phénotypage numérique suggère de collecter passivement les données des patients, ce qui permet une surveillance non intrusive et continue des états comportementaux et mentaux et, au final, de révéler des informations cliniquement pertinentes. De même, les « marqueurs numériques » sont des indicateurs de maladie obtenus par voie numérique qui peuvent être utilisés pour définir un phénotype numérique [2].

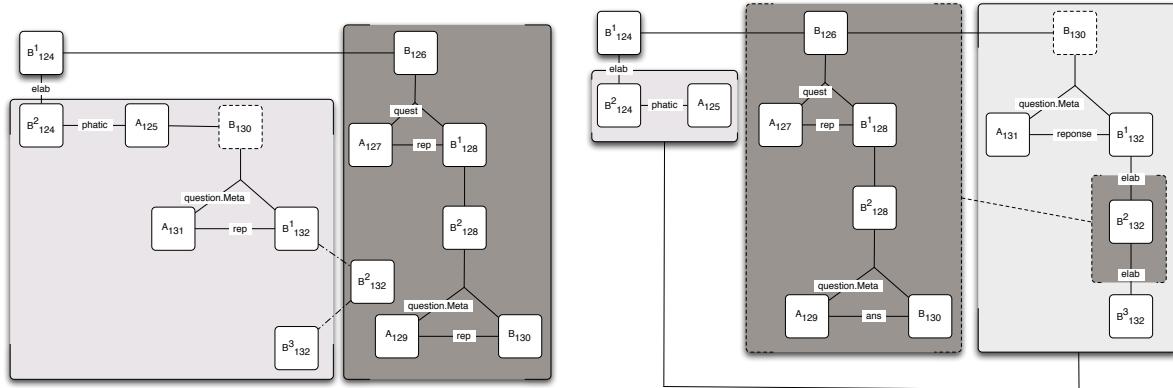


Figure 1: Application du modèle du discours aux entretiens pathologiques.

À partir du recueil de données, nous développons une entreprise de formalisation psycholinguistique et sémantico-formelle engagée dans le but d'identifier des signes diagnostiques aussi bien que prodromiques dans le discours (action Exploratoire ODiM et programme MEPHESTO). La réussite d'un tel projet, étape préalable nécessaire à la construction d'outils d'aide au diagnostic informatisé, nécessite l'extension de la méthodologie d'analyse mise au point antérieurement par les participants du projet afin d'augmenter fortement la capacité du modèle initial à identifier les séquences porteuses de discontinuités symptomatiques propres à la schizophrénie et à d'autres maladies mentales [1, 7]. Nous illustrons dans la Figure 1 le type de modèle que nous produisons pour le discours. Ce travail d'extension est réalisé au travers d'une expertise

scientifique, et également par des analyses systématiques grâce aux outils automatiques. Les théories utilisées comme fondement ont été élaborées indépendamment des singularités du discours « pathologique ». Il convient maintenant de modéliser le couplage « configuration discursive-expression interactionnelle d'un symptôme psychopathologique » en élaborant une sémantique formelle originale et adaptée. Pour cela nous mobiliserons les stratégies récemment développées autour de la théorie des continuations et du  $\lambda$ -calcul. Il convient aussi de révéler les processus cognitifs, en particulier neuropsychologiques (fonctions exécutives) sous-jacents. Cette double perspective centrée sur l'identification de signes pathognomoniques des pathologies mentales en termes de troubles du discours, à la fois « cognitiviste » et « sémantico-formelle », constitue un renouvellement de paradigme méthodologique pour les participants du projet. Elle n'a par ailleurs jamais été tentée jusqu'aujourd'hui par d'autres équipes. Un axe de ce projet consiste ainsi à explorer le caractère heuristique du modèle d'analyse des discontinuités de l'interaction conversationnelle pré-élaboré. Son développement en matière de repérage des troubles du discours nécessite l'élaboration d'un formalisme idoine qui prenne en compte des mécanismes de construction de l'interaction d'un niveau de complexité que les théories actuelles ne prennent pas en charge.

Ces dernières années, les méthodes de reconnaissance du comportement basées sur l'intelligence artificielle (c'est-à-dire l'apprentissage automatique ou profond) sont devenues de plus en plus efficaces dans une variété de tâches, y compris la classification des actions, le langage corporel et les gestes, l'estimation du regard, la détection du contact visuel, la détection des mouvements de regard et la fréquence saccadique, les unités d'action du visage, ainsi que l'affect extrait de modalités uniques ou multiples.

Un nombre croissant d'approches utilisent ces progrès dans la détection du comportement humain pour analyser les données d'interaction clinique (par exemple, les séances de thérapie et de suivi clinique) et les caractéristiques linguistiques et paralinguistiques de la parole. Comme les troubles psychiatriques (TDM, bipolarité, schizophrénie) ont un impact sur la qualité des interactions sociales, l'accent est mis sur l'étude de ces dynamiques comportementales quantifiables dans les interactions sociales de la vie réelle au niveau de la dyade plutôt que sur le seul comportement individuel.

L'analyse automatique de la parole et de l'audio donne des résultats prometteurs pour la détection des états affectifs chez les patients souffrant de dépression, de troubles bipolaires, de schizophrénie ou de démence. Certaines technologies mobiles ont démontré la faisabilité du suivi de la dépression qui pourrait alimenter des modèles de prédiction des rechutes. Dans le cas de la schizophrénie, les données passives des smartphones ont été utilisées pour examiner le comportement des utilisateurs et prédire les rechutes cliniques. L'étude CrossCheck a révélé de fortes associations et un pouvoir prédictif entre les symptômes autodéclarés et les données actives et passives.

Bien que ces premiers résultats soient prometteurs, ces recherches doivent être accélérées par le développement de plateformes technologiques de phénotypage numérique axées sur l'évolutivité et l'équité, par l'établissement de référentiels de données longitudinales partagées et par l'encouragement de collaborations multidisciplinaires entre les intervenants cliniques, notamment les patients, les informaticiens et les chercheurs [5].

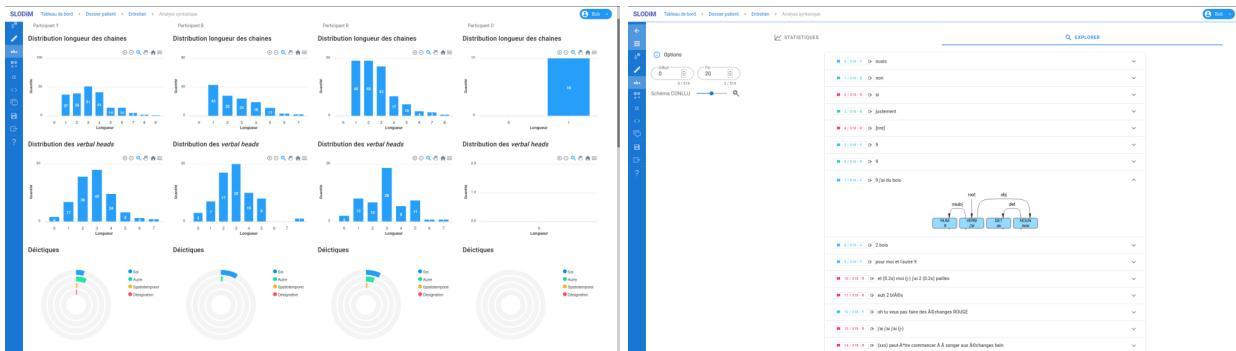


Figure 2: Exemple d'analyse automatique produite par l'outil SloDiM.

L'objectif est d'outiller les cliniciens dans leur prise de décision diagnostique, ainsi que dans le suivi au long cours de l'état de santé de leurs patients. Nous avons pris en charge ces questions dans les projets MePheSTO et ODIM. Si l'objectif général est de proposer un cadre méthodologique permettant de définir et

d’appréhender des indices prodromiques ou diagnostiques associés à la psychose (schizophrénie, bipolarité), nous souhaitons aussi outiller ces approches par le développement de logiciels permettant d’identifier automatiquement ces indices, tant au plan des symptômes des patients repérables au niveau du discours qu’au niveau de leur comportement langagier. Un premier aspect est de mobiliser les outils au niveau de l’état de l’art pour le français afin de produire une analyse systématique de la production langagière, ce qui permettra d'affiner le développement du modèle. Un second aspect est de développer des outils d’analyse automatique pour des niveaux linguistiques plus complexes dans les dysfonctionnements langagiers identifiés. Nous avons développé un premier outil, réalisant les analyses limitées aux parties du discours (POS) et disfluences, étendues à l’analyse syntaxique comme visible dans la Figure 2. Cela nous a permis de discuter des spécificités des pratiques langagières. Il convient maintenant de reprendre ces développements pour proposer des analyses sur les niveaux plus conceptuels. Nous devons développer une chaîne de traitement et d’analyse, afin de se concentrer sur le nœud sémantico-pragmatique (cf partie précédente). Il existe peu de ressources ou d’outils en la matière. D’un point de vue opérationnel, il s’agit de disposer d’analyse sémantique et discursive (ou dialogique) de l’ensemble des entretiens. Actuellement, nous concevons les représentations logiques (sémantiques). Le développement de cet outil revêt plusieurs aspects. Il s’agit d’approfondir la qualité d’analyse des corpus dont nous disposons et dans le même temps de disposer d’une chaîne de traitement robuste.

## References

- [1] Maxime Amblard, Michel Musiol, and Manuel Rebuschi. *(In)coherence of Discourse. Formal and conceptual issues of language*. Language, Cognition and Mind (Chungmin Lee “Editor Springer book series). Springer, June 2021.
- [2] A Andrea, A Agulia, G Serafini, and M Amore. Digital biomarkers and digital phenotyping in mental health care and prevention. *European Journal of Public Health*, 30(Supplement\_5):cka165–1080, 2020.
- [3] Armen C Arevian, Daniel Bone, Nikolaos Malandrakis, Victor R Martinez, Kenneth B Wells, David J Miklowitz, and Shrikanth Narayanan. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS one*, 15(1):e0225695, 2020.
- [4] Elisa Brietzke, Emily R Hawken, Maia Idzikowski, Janice Pong, Sidney H Kennedy, and Claudio N Soares. Integrating digital phenotyping in clinical characterization of individuals with mood disorders. *Neuroscience & Biobehavioral Reviews*, 104:223–230, 2019.
- [5] Kit Huckvale, Svetha Venkatesh, and Helen Christensen. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ digital medicine*, 2(1):1–11, 2019.
- [6] Thomas R Insel. Digital phenotyping: technology for a new science of behavior. *Jama*, 318(13):1215–1216, 2017.
- [7] Michel Musiol, Manuel Rebuschi, Samuel Buchel, Amandine Lecomte, Philippe de Groote, and Maxime Amblard. Le problème de l’analyse des troubles de la pensée dans le discours avec la personne schizophrène : proposition méthodologique. *L’Evolution Psychiatrique*, 86(2), 2022.