Demande n°3869 - ARED

Acronyme :STP

Identification

Identification

N°	3869
Année	2025
Statut	Brouillon
Date de création	28/01/2025
Date de modification	29/01/2025
Date de soumission	29/01/2025
Acronyme	STP
Intitulé du projet	Stéganographie Textuelle par Paraphrase
Dispositif	ARED

Priorité régionale

Stratégie régionale de recherche et d’innovation (S3)

Stratégie régionale de recherche et d’innovation (S3)

Domaine d'innovation stratégique principal	Économie numérique sécurisée et responsable
Levier thématique associé au DIS principal	Cybersécurité
Domaine d'innovation statégique secondaire	Économie numérique sécurisée et responsable
Levier thématique associé au DIS secondaire	Données et intelligence

Organisme de tutelle

Organisme de tutelle

Etablissement	UNIVERSITE RENNES 1
Commune	RENNES
Civilité	Monsieur

Nom	Le Gouguec
Prénom	Fabrice

Unité de recherche

Unité de recherche

Nom de l'unité	IRISA
Civilité	Monsieur
Nom	Gravier
Prénom	Guillaume

Données complémentaires de l'unité de recherche

Libellé	Valeur
Statut et numéro de l'unité de recherche (UMR, USR, EA)	UMR 6074
Le cas échéant, nom de l'équipe de recherche	Expression
Le cas échéant, nom du responsable de l'équipe de recherche	Pierre-François Marteau (Professeur, Univ. Bretagne Sud)
Activités de recherche et points forts de l'équipe	Expression cible l'étude du langage humain véhiculé par différents canaux : le geste, la parole et le texte. L'équipe s'est positionnée depuis plusieurs années comme spécialiste du traitement et de la génération du langage, notamment écrit.
Si implication dans un/plusieurs programmes européens ou internationaux, préciser	

Co-directeur de thèse

Description

Description du projet

Libellé	Valeur
Résumé synthétique du projet (2000 caractères max)	La stéganographie est une technique qui consiste à dissimuler une information sensible dans un média de couverture à l'apparence anodine, ici un texte. Contrairement au tatouage de documents, qui vise à rendre l'information inamovible des données de couverture, la stéganographie cherche à rendre cette information indétectable tout en maximisant la quantité d'information dissimulée dans un document de taille donnée. Les approches récentes en stéganographie textuelle s'appuient souvent sur la génération de textes de couverture respectant une thématique prédéfinie. Cependant, ces textes générés peuvent manquer de crédibilité dans le contexte d'utilisation. De plus, la quantité d'informations qui doit être pré-partagée entre l'émetteur et le récipiendaire est parfois disproportionnée (modèle informatique, thématique du message de couverture, etc.), ce qui peut rendre les méthodes proposées irréalistes. Par ailleurs, la production de paraphrases a considérablement progressé ces dernières années, notamment grâce aux avancées des grands modèles de langage (LLM). Si la réécriture de textes en conservant leur sens est de plus en plus performante, le contrôle précis de cette réécriture pour des tâches spécifiques reste un verrou scientifique difficile. La stéganographie textuelle représente une application ambitieuse dans ce domaine, car elle impose un critère strict au texte généré : le message secret doit être retrouvé par le

Libellé	Valeur
	au texte généré : le message secret doit être retrouvé par le récipiendaire, et uniquement par lui. L'objectif de cette thèse est donc d'explorer la possibilité de modifier un texte existant pour y dissimuler un message tout en préservant son sens original. Des résultats collatéraux sur la détection de stéganographie textuelle sont également attendus.
Caractère structurant/stratégique du projet (descriptif/ex. : art.5 Rglmt ARED 2025 -2000 car. max)	Ce sujet s'inscrit dans la continuité des travaux sur la production automatique et contrôlée de paraphrases au sein de l'équipe Expression. Il est en synergie avec la thèse DGA de M. Lemesle, actuellement en deuxième année. Au sein du pôle de recherche en informatique de Lannion, ce sujet permettrait d'amorcer un pont entre l'équipe Expression (IA, génération de texte) et l'équipe Sautern (cybersécurité, réseaux). Enfin, le projet s'inscrit dans la volonté de renforcer le nombre d'HDR sur le site de Lannion, puisque le porteur, qui a déjà entrepris les démarches pour s'inscrire en HDR, s'engage à la défendre avant le début de la thèse.
Porteur du projet : date d'obtention de l'HDR ou dérogation	
Collaborations scientifiques (nature/pays) et partenariat socio-économique envisagé (3000 car. max.)	Les travaux sur la génération de textes intéressent de nombreux acteurs déjà en interaction avec le porteur du projet : Orange Innovation, avec, entre autres, une thèse co-encadrée soutenue en 2022. La DGA, via une thèse sur la paraphrase contrôlée en minimisant les données d'apprentissage. Plusieurs PME/startups locales ou nationales qui ont sollicité l'équipe sur ce sujet (SeeLab, Bloom, Anemon, etc.). L'axe de la stéganographie pourrait également permettre de renforcer des collaborations avec des acteurs du domaine cyber, tels que la DGA ou Thales.
Si DIS identifié, préciser l'adéquation du projet avec le DIS (2000 car. maximum)	Le projet est en adéquation avec le DIS D3 "Économie numérique sécurisée et responsable", en particulier via les leviers "Cybersécurité" et "Données et intelligence artificielle". En effet, ces travaux s'intéressent à la production automatique de texte via des grands modèles de langage (LLM). Une des questions sous-jacentes est de pouvoir contrôler précisément ce qu'ils produisent tout en conservant le sens du texte original. Dans un monde où l'usage des LLM se multiplie (ChatGPT, DeepSeek, Mistral), la maîtrise de leur production devient un enjeu crucial. L'application à la stéganographie textuelle est particulièrement intéressante, au-delà des considérations scientifiques, car ce domaine reste relativement peu étudié jusqu'à présent. D'un point de vue cybersécurité, la maîtrise de l'information au sein des organisations est un enjeu essentiel. La possibilité d'exfiltrer des données via un texte cohérent, sans utiliser les propriétés du support (encodage, caractères invisibles, etc.), permettrait de contourner les mécanismes de contrôle, aussi bien humains qu'automatiques. Ainsi, en étudiant les stratégies de stéganographie textuelle, la question de leur détection sera bien évidemment considérée.
Si priorité régionale cochée, préciser en quoi le projet s'y rattache effectivement (2000 car.max.)	
Si priorité régionale "Politique publique Région", préciser (voir liste au point 4, Règlement 2025)	NA
Cotutelle financière internationale (oui/non)	Non
Nom(s) du(des) cofinanceur(s) envisagé(s) ou confirmé(s)	- CMA CyberSkill4All - Lannion-Trégor Communauté - DGA AMI
Si cofinancement industriel, justifier le choix du dispositif ARED au lieu du dispositif CIFRE	

État du cofinancement (indiquer "acquis" ou "demandé")	- Lannion-Trégor Communauté (acquis) - CMA CyberSkill4All (demandé)
Si demandé, date prévisionnelle de réponse du cofinanceur	- CyberSkill4All : Juin 2025
Si cofinancement refusé, autre(s) source(s) de cofinancement identifiée(s)	
Si reversement prévu, préciser (partenaire, montant)	- Lannion-Trégor Communauté : 50%

Candidat

Collaboration internationale

Collaboration internationale

Pays	
Etablissement partenaire à l'étranger	

Pièces jointes

Date de mi	Type de pièce	Commentaire	Document	Valide	Commentaire
28/01/2025	ARED - Dépôt - CV du porteur	Obligatoire - Le CV doit préciser si le porteur est titulaire de l'HDR et, le cas échéant, depuis quand, et mentionner la liste des doctorant·e·s déjà encadré·e·s et des publications associées.	CV-JonathanChevelu.pdf	Non traité	
28/01/2025	ARED - Dépôt - Avis directeur·rice d'unité	Obligatoire - Voir le modèle à suivre dans l'onglet "Documents à télécharger"		Non traité	
28/01/2025	ARED - Dépôt - Justificatif cofinancement	Facultatif - Dépôt de demande, justificatif d'obtention, courrier du cofinanceur pressenti...		Non traité	
28/01/2025	ARED - Dépôt - Lettre partenaire étranger	Facultatif - Lettre d'engagement de l'établissement partenaire étranger en cas de cotutelle internationale		Non traité	
28/01/2025	ARED - Dépôt - Pièce complémentaire	Facultatif		Non traité	
28/01/2025	ARED - Dépôt - Pièce complémentaire	Facultatif		Non traité	
28/01/2025	ARED - Instruction - Lettre de saisine	Lettre de l'établissement à transmettre à la Région (utiliser le modèle disponible dans la rubrique "Documents à télécharger")		Non traité	
28/01/2025	ARED - Versement - Convention de partenariat	Le cas échéant (si autorisation de reversement)		Non traité	
28/01/2025	ARED - Versement 2 - CV doctorant.e	Pour versement tranche 2		Non traité	
28/01/2025	ARED - Versement 2 - Copie du contrat doctoral (ou CDD ou attestation))	Pour versement tranche 2		Non traité	

28/01/2025	ARED - Versement 2 - Attestation bon déroulement travaux année 1	Pour versement tranche 2		Non traité	
28/01/2025	ARED - Versement 2 - Bilan financier année 1	Pour versement tranche 2 ARED		Non traité	
28/01/2025	ARED - Versement 3 - Attestation bon déroulement travaux année 2	Pour versement 50% tranche 3 ARED		Non traité	
28/01/2025	ARED - Versement 3 - Bilan financier année 2	Pour versement 50% tranche 3 ARED		Non traité	
28/01/2025	ARED - Versement 4 - Bilan financier année 3	Pour versement solde année 3 ARED		Non traité	
28/01/2025	ARED - Versement 4 - Certificat administratif précisant statut thèse	Pour versement du solde de la tranche 3		Non traité	

Bureau 210I
IRISA / ENSSAT - Univ Rennes
6 rue de Kerampont - BP 80518
F-22305 Lannion Cedex, France

jonathan.chevelu@irisa.fr
+33 2 96 46 91 84

CHEVELU
Jonathan

41 ans
Nationalité française

Poste actuel

2017- (7 ans)	**Maître de conférences** Université de Rennes IRISA **Lannion**	– enseignement en informatique dans l'école d'ingénieur ENSSAT ; – recherche en traitement automatique de la parole. Production automatique de séquences (texte, voix) ;

Expériences professionnelles

2015-2016 (1 ans)	**ATER en Informatique** Université de Rennes1 IRISA **Lannion**	– enseignement : algorithmique, cybersécurité, structure de données, web ; – recherche : synthèse vocale, construction optimale de corpus.
2013-2015 (2 ans)	**Ingénieur de recherche** Université de Rennes1 IRISA **Lannion**	– création de contenus dynamiques pour l'apprentissage de l'écrit par synthèse vocale ; – projet ANR Phorevox ; – synthèse vocale, dictée automatique, exercices phonétiques.
2011-2012 (6 mois)	**Ingénieur R&D** Critéo **Paris**	– publicité ciblée et reciblée sur internet ; – une solution de ciblage publicitaire par mots-clefs ; – étude de la classification de texte en catégories commerciales.
2011 (8 mois)	**Ingénieur de recherche** Twenga **Paris**	– moteur de shopping sur internet ; – production automatique de descriptions et de titres ; – analyses, prédictions et optimisations du trafic et du référencement.
2007-2010 (3 ans)	**Doctorant** Université de Caen Basse-Normandie **Orange Labs, Lannion**	– production de paraphrases pour le dialogue humain-machine ; – création d'un outil d'optimisation par méthode de Monte-Carlo ; – ajout du multilinguisme dans l'agrégateur de news d'Orange ;

Formation et diplômes

2007–2011	**Doctorat** de l'Université de Caen Basse-Normandie en informatique et applications.
2004–2007	**Magistère de l'École Normale Supérieure** de Cachan en informatique et télécommunications.
2005–2007	**Master recherche** de l'Université de Rennes en informatique mention Bien.

Enseignement

2013-	**Niveau L1 à Master** ENSSAT **Lannion**	– Traitement automatique des langues et de la parole, apprentissage automatique, systèmes d'exploitation, cybersécurité, développement informatique ; – Service à 192h équivalent TD par an (hors heures complémentaires).
2018-2022	**Formation continue** ENSSAT **Lannion**	– Participation à la construction et l'enseignement pour la formation « Référent cybersécurité » ; – Labellisée SecNumedu par l'ANSSI.

Responsabilités scientifiques et para-pédagogiques

2024-	**Responsable élu du pôle informatique** de l'école d'ingénieur ENSSAT.
2020-	**Membre élu du conseil scientifique** de l'ENSSAT.
2018-	**Responsable de la première année informatique** de l'ENSSAT.
2015-	**Co-fondateur, co-organisateur et responsable technique** du challenge de programmation pour lycéennes **Ada Lovelace** (https://adalovelace.ledantec-numerique.fr/).
2020-2022	**Membre élu de la commission du personnel** de l'ENSSAT.
2010-2012	**Membre élu au CA** de l'Association pour le Traitement Automatique des Langues (**ATALA**).

Activité de recherche en bref

Thématiques	**Production automatique de paraphrase, syntèse vocale, traitement automatique des langues, clonage vocal, apprentissage automatique.**
Publications	**1 journal international [1], 1 journal francophone [2], 26 conférences internationales [3–28], 3 ateliers internationaux [29–31], 1 chapitre de livre [32], 8 conférences francophones [33–40], 3 dépôts à l'APP [41–43], 2 démonstrations [44, 45].**
Encadrements	**5 co-encadrements de thèses** (3 soutenues et 2 en cours) **; 2 post-doctorants ; 2 masters recherche.**
Projets	**CRE SeeLab**, **ANR Text2Kids**, **ANR JCJC Tremolo**, **ANR Phorevox**.
Relecture	**Interspeech, ICASSP, Coling, LREC, ANR.**

Encadrement de thèses

2023-	**MAS Pauline** CIFRE Voxygen	– *Modulation multilingue d'éléments prosodiques en synthèse vocale.* ▷ *Direction 50% (par dérogation à l'HDR), Encadrement à 25%.*
2023-	**LEMESLE Quentin** DGA / CDO UR1	– *Zero-shot learning pour la production de paraphrases sous contraintes.* ▷ *Encadrement à 34% ; Publications : [3, 33].*
2020-2024	**WADOUX Lily** CDO UR1	– *Impact de la qualité et du contenu des corpus sur le clonage de voix : Vers une application en français aux voix pathologiques.* – Soutenue le 21/11/2024. ▷ *Encadrement à 50% ; Publications : [4, 5].*
2018-2022	**FABRE Betty** CIFRE Orange Labs	– *Search-Based And Supervised Text Generation.* – Soutenue le 16/9/2022. ▷ *Encadrement à 33% ; Publications : [6, 30].*
2017-2020	**SHAMSI Meysam** ARED / CD22	– *Script optimization for TTS voice corpus design in audio-book generation.* – Soutenue le 16/10/2020. ▷ *Encadrement à 33% ; Publications : [9–12].*

Projets et travaux collaboratifs

2025	**SeeLab** CRE / BPI	– Opportunité d'utiliser de méthode de production automatique de texte pour améliorer de la production d'images personnalisées. – Projet d'un an avec la PME SeeLab financé par la BPI. ▷ *Porteur du projet pour l'IRISA. Budget ∼25 000€.*
2024-2029	**CyberSkills4All** CMA	– Compétences en cyber sécurité pour tous. – UBS ; Rennes School of Business ; ENSTA Bretagne ; ENI Brest ; Univ Rennes, FUN ; PEC ; ORANGE SA ; CMQe Bretagne. ▷ *Participant. Budget ∼23M€.*
2019-2024	**Text2Kids** ANR PRCE	– Étudie des caractéristiques linguistiques propices à une compréhension optimale d'un contenu informationnel par des enfants [7]. – Modyco ; IRISA ; Synapse Developpement ; QWANT. ▷ *Responsable Work Package. Budget ∼650 000€.*
2017-2020	**Tremolo** ANR JCJC	– Étudier et caractériser l'emploi des registres de langue dans des textes écrits [13, 36]. – IRISA ; Modico. ▷ *Responsable Work Package. Budget ∼250 000€.*
2015-2017	**Blizzard Challenge**	– Participation au challenge international de synthèse de la parole avec le moteur de l'équipe Expression [42]. ▷ 2015 : synthèse pour 5 langues indiennes [19] ; ▷ 2016 et 2017 : synthèse de livre audio en anglais [14, 15].
2013-2015	**Phorevox** ANR	– Apprentissage du français écrit par l'usage de technologies vocales [44]. – IRISA ; Zeugmo S.A.S ; CREAD ; LLF ; Voxygen. ▷ *Ingénieur de recherche, aide à la coordination. Budget ∼450 000€.*

Thématiques de recherche

Mes travaux s'intéressent principalement à la production/sélection automatique et contrôlée de textes.
Ces textes peuvent servir à alimenter de façon optimale un système de traitement automatique, par exemple pour de la synthèse de la parole [1, 4, 11, 25]. Ces recherches m'ont aussi conduit à étudier la synthèse de la parole de manière concomitante [10, 18, 19].
Les textes produits peuvent aussi être directement adaptés en fonction des utilisateurs (âge [7], registre de langue [13], pathologie vocale [5]) ou en fonction d'une tâche (dialogue [21], dictée [44]). Je me suis particulièrement intéressé à la notion de paraphrase, c'est-à-dire de conservation du sens original du message[6, 23].
De plus, les questions d'évaluations perceptives et automatiques des productions ont notamment été étudiées [3, 17, 35, 39].

Publications et communications

Journal international avec comité lecture

[1] N. BARBOT, O. BOËFFARD, J. CHEVELU et A. DELHAY : Large Linguistic Corpus Reduction with SCP algorithms. *Computational Linguistics*, 41(3) :30, sept. 2015. URL https://inria.hal.science/hal-01135089.

Journal francophone avec comité lecture

[2] A. SINI, L. WADOUX, A. PERQUIN, G. VIDAL, D. GUENNEC, D. LOLIVE, P. ALAIN, N. BARBOT, J. CHEVELU et A. DELHAY : Techniques de synthèse vocale neuronale à l'épreuve des données d'apprentissage non dédiées : les livres audio amateurs en français. *Revue TAL : traitement automatique des langues*, 63(2) :41–65, juin 2023. URL https://inria.hal.science/hal-04139701.

Conférences internationales avec comité lecture

[3] Q. LEMESLE, J. CHEVELU, P. MARTIN, D. LOLIVE, A. DELHAY et N. BARBOT : Paraphrase Generation Evaluation Powered by an LLM : A Semantic Metric, Not a Lexical One. *In The 31st International Conference on Computational Linguistics (COLING)*, Abu Dhabi, United Arab Emirates, jan. 2025. International Committee on Computational Linguistics. URL https://hal.science/hal-04849291.

[4] L. WADOUX, N. BARBOT, J. CHEVELU et D. LOLIVE : Voice Cloning for Voice Disorders : Impact of Phonetic Content. *In TSD 2023 : Text, Speech, and Dialogue*, vol. 14102 de *Lecture Notes in Computer Science*, p. 293–303, Pilsen, Czech Republic, sept. 2023. Springer Nature Switzerland. URL https://inria.hal.science/hal-04223342.

[5] L. WADOUX, N. BARBOT, D. LOLIVE et J. CHEVELU : Voice Cloning Applied to Voice Disorders : a Study of Extreme Phonetic Content in Speaker Embeddings. *In 35th Canadian Conference on Artificial Intelligence*, Toronto, Canada, mai 2022. URL https://inria.hal.science/hal-03697484.

[6] B. FABRE, T. URVOY, J. CHEVELU et D. LOLIVE : Neural-Driven Search-Based Paraphrase Generation. *In 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, Kiev, Ukraine, avr. 2021. URL https://hal.science/hal-03540926.

[7] R. RAHMAN, G. LECORVÉ, A. ÉTIENNE, D. BATTISTELLI, N. BÉCHET et J. CHEVELU : Mama/Papa, Is this Text for Me ? *In The 28th International Conference on Computational Linguistics (COLING)*, p. 6296–6301, Virtuel, Spain, déc. 2020. International Committee on Computational Linguistics. URL https://hal.science/hal-03112352.

[8] M. SHAMSI, J. CHEVELU, N. BARBOT et D. LOLIVE : Corpus design for expressive speech : impact of the utterance length. *In 10th International Conference on Speech Prosody 2020*, p. 955–959, Tokyo, Japan, mai 2020. ISCA. URL https://hal.science/hal-02874005.

[9] M. SHAMSI, N. BARBOT, D. LOLIVE et J. CHEVELU : Mixing Synthetic and Recorded Signals for Audio-Book Generation. *In Speech and Computer (SPECOM) 2020*, p. 479–489, sept. 2020. URL https://inria.hal.science/hal-02974346.

[10] M. SHAMSI, D. LOLIVE, N. BARBOT et J. CHEVELU : Script Selection using Convolutional Auto-encoder for TTS Speech Corpus. *In International Conference on Speech and Computer (SPECOM)*, Istanbul, Turkey, août 2019. Springer. URL https://inria.hal.science/hal-02169465.

[11] M. SHAMSI, D. LOLIVE, N. BARBOT et J. CHEVELU : Corpus Design using Convolutional Auto-Encoder Embeddings for Audio-Book Synthesis. *In Interspeech*, Graz, Austria, sept. 2019. URL https://inria.hal.science/hal-02159568.

[12] M. SHAMSI, D. LOLIVE, N. BARBOT et J. CHEVELU : Investigating the relation between voice corpus design and hybrid synthesis under reduction constraint. *In 7th International Conference on Statistical Language and Speech Processing (SLSP)*, Ljubljana, Slovenia, oct. 2019. URL https://inria.hal.science/hal-02178406.

[13] G. LECORVÉ, H. AYATS, B. FOURNIER, J. MEKKI, J. CHEVELU, D. BATTISTELLI et N. BÉCHET : Towards the Automatic Processing of Language Registers : Semi-supervisedly Built Corpus and Classifier for French. *In International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, vol. 13451 de *Lecture Notes in Computer Science*, p. 480–492, La Rochelle, France, avr. 2019. Springer Nature Switzerland. URL https://hal.science/hal-02064694.

[14] D. LOLIVE, P. ALAIN, N. BARBOT, J. CHEVELU, G. LECORVÉ, C. J. SIMON et M. TAHON : The IRISA Text-To-Speech System for the Blizzard Challenge 2017. *In Blizzard Challenge*, Stockholm, Sweden, août 2017. URL https://inria.hal.science/hal-01662361.

[15] P. ALAIN, J. CHEVELU, D. GUENNEC, G. LECORVÉ et D. LOLIVE : The IRISA Text-To-Speech System for the Blizzard Challenge 2016. *In Blizzard Challenge 2016 workshop*, Cupertino, United States, sept. 2016. URL https://inria.hal.science/hal-01375897.

[16] J. CHEVELU et D. LOLIVE : Do not build your TTS training corpus randomly. *In Proceedings of the 2015 European Signal Processing Conference (EUSIPCO)*, Nice, France, août 2015. European Association for Signal Processing (EURASIP). URL https://inria.hal.science/hal-01199083.

[17] J. CHEVELU, D. LOLIVE, S. LE MAGUER et D. GUENNEC : How to Compare TTS Systems : A New Subjective Evaluation Methodology Focusing on the Differences. *In Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech)*, Dresden, Allemagne, sept. 2015. International Speech Communication Association (ISCA). URL https://inria.hal.science/hal-01199082.

[18] D. GUENNEC, J. CHEVELU et D. LOLIVE : Defining a Global Adaptive Duration Target Cost for Unit Selection Speech Synthesis. *In International Conference on Text, Speech and Dialogue (TSD)*, p. 157–165, Plzeň, République Tchèque, sept. 2015. Springer. URL https://inria.hal.science/hal-01188686.

[19] P. ALAIN, J. CHEVELU, D. GUENNEC, G. LECORVÉ et D. LOLIVE : The IRISA Text-To-Speech System for the Blizzard Challenge 2015. *In Blizzard Challenge 2015 Workshop*, p. 4 p., 2 columns, Berlin, Allemagne, sept. 2015. URL https://inria.hal.science/hal-01196168.

[20] J. CHEVELU, G. LECORVÉ et D. LOLIVE : ROOTS : a toolkit for easy, fast and consistent processing of large sequential annotated data collections. *In Proceedings of the 9th International Language Resources and Evaluation Conference (LREC)*, Reykjavik, Island, mai 2014. European Language Resources Association (ELRA). URL https://inria.hal.science/hal-00974628.

[21] G. PUTOIS, J. CHEVELU et C. BOIDIN : Paraphrase generation to improve Text-To-Speech Synthesis. *In Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech)*, Tokyo, Japon, sept. 2010. International Speech Communication Association (ISCA). URL http://www.interspeech2010.org/program/session.php?id=2230.

[22] J. CHEVELU, G. PUTOIS et Y. LEPAGE : The True Score of Statistical Paraphrase Generation. *In Proceedings of the 23rd Conference on Computational Linguistics (Coling) : Posters*, p. 144–152, ékin, Chine, août 2010. Coling 2010 Organizing Committee. URL https://hal.science/hal-01066835.

[23] J. CHEVELU, T. LAVERGNE, Y. LEPAGE et T. MOUDENC : Introduction of a new paraphrase generation tool based on Monte-Carlo sampling. *In Proceedings of the ACL-IJCNLP Conference Short Papers*, p. 249–252, Singapore, Singapore, août 2009. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P/P09/P09-2063.

[24] J. CHEVELU, T. LAVERGNE, Y. LEPAGE et T. MOUDENC : Transformation Rules and Monte-Carlo Sampling : a Different Approach for Statistical Paraphrase Generation. *In* H. KAMEDA, M. TOKUHISA, S. OHNO, M. SUZUKI et J. SJÖBERGH, éds : *Proceedings of the Conference of the Pacific Association for Computational Linguistics (PACLING)*, p. 230–235, Sapporo, Japon, sept. 2009. URL http://sig.media.eng.hokudai.ac.jp/pacling2009/index.html.

[25] C. BOIDIN, V. RIESER, L. VAN DER PLAS, O. LEMON et J. CHEVELU : Predicting how it sounds : Re-ranking dialogue prompts based on TTS quality for adaptive Spoken Dialogue Systems. *In Proceedings of the Interspeech Special Session : Machine Learning for Adaptivity in Spoken Dialogue*, p. 2487–2490, Brighton, UK, sept. 2009. International Speech Communication Association (ISCA). URL http://www.interspeech2009.org/conference/programme/session.php?id=6510.

[26] J. CHEVELU, N. BARBOT, O. BOËFFARD et A. DELHAY : Comparing set-covering strategies for optimal corpus design. *In* N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODJIK, S. PIPERIDIS et D. TAPIAS, éds : *Proceedings of the 6th International Language Resources and Evaluation Conference (LREC)*, Marrakech, Maroc, mai 2008. European Language Resources Association (ELRA). ISBN 2-9517408-4-0. URL https://hal.science/hal-04649815.

[27] K. YOUSFI, C. AMBROISE, J.-P. COCQUEREZ et J. CHEVELU : Supervised Learning for Guiding Hierarchy Construction : Application to Osteo-Articular Medical Images Database. *In Procedings of the 18th International Conference on Pattern Recognition (ICPR)*, vol. 1, p. 484–487, Hong Kong, Chine, sept. 2006. IEEE. ISBN 0-7695-2521-0. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1698937.

[28] K. YOUSFI, C. AMBROISE, J.-P. COCQUEREZ et J. CHEVELU : Driving Hierarchy Construction via Supervised Learning : Application to Osteo-Articular Medical Images Database. *In Procedings of the International Conference on Image Processing (ICIP)*, p. 2433–2436, Atlanta, USA, oct. 2006. IEEE. ISBN 1-4244-0480-0. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4107059.

Ateliers internationaux avec comité lecture

[29] N. DEHGHANI, H. HAJIPOOR, J. CHEVELU et G. LECORVÉ : User-oriented paraphrase generation with multiple types of constraints. *In NeurIPS workshop on Controllable Generation for Modeling in Language and Vision (CtrlGen)*, 2021. URL https://ctrlgenworkshop.github.io/camready/11/CameraReady/CtrlGen_NeurIPS_2021.pdf.

[30] B. FABRE, T. URVOY, J. CHEVELU et D. LOLIVE : Neural-Driven Multi-criteria Tree Search for Paraphrase Generation. *In Learning Meets Combinatorial Algorithms (LMCA) Workshop at NeurIPS 2020*, online, France, déc. 2020. URL https://inria.hal.science/hal-03127865.

[31] J. CHEVELU, N. BARBOT, O. BOËFFARD et A. DELHAY : Lagrangian Relaxation for Optimal Corpus Design. *In Proceedings of the 6th ISCA Tutorial and Research Workshop on Speech Synthesis (SSW)*, p. 211–216, Bonn, Allemagne, août 2007. ISCA. URL https://hal.science/hal-04649887.

Chapitre d'ouvrage

[32] G. J. McLachlan, J. Chevelu et J. X. Zhu : Correcting for selection bias via cross-validation in the classification of microarray data. *Beyond Parametrics in Interdisciplinary Research : Festschrift in Honor of Professor Pranab K. Sen*, 1 :364–376, 2008. URL http://www.projecteuclid.org/DPubS/Repository/1.0/Disseminate?handle=euclid.imsc/1207058286&view=body&content-type=pdf_1.

Conférences francophones avec comité lecture

[33] Q. Lemesle, J. Chevelu, D. Lolive, A. Delhay et P. Martin : ParaPLUIE - une mesure automatique d'évaluation de la qualité sémantique des systèmes de paraphrases. *In* Balaguer, Mathieu, Bendahman, Nihed, Ho-Dac, Lydia-Mai, Mauclair, Julie, Moreno, J. G, Pinquier et Julien, éds : *Actes de la 31{è}me Conf{é}rence sur le Traitement Automatique des Langues Naturelles (TALN)*, vol. 1 : articles longs et prises de position, p. 605–616, Toulouse, France, juil. 2024. ATALA & AFPC. URL https://inria.hal.science/hal-04623043.

[34] L. Wadoux, N. Barbot, J. Chevelu et D. Lolive : Impact du contenu phonétique sur les plongements de locuteurs pour le clonage de voix : vers l'application aux pathologies vocales. *In XXXIVe Journées d'études sur la Parole – JEP 2022*, p. 814–822, Noirmoutier-en-l'Ile, France, juin 2022. ISCA. URL https://hal.science/hal-03975729.

[35] C. Fayet, A. Blond, G. Coulombel, C. Simon, D. Lolive, G. Lecorvé, J. Chevelu et S. Le Maguer : FlexEval, création de sites web légers pour des campagnes de tests perceptifs multimédias. *In* C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla et S. Schneider, éds : *6e conférence conjointe Journées d'études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RéCITAL, 22e édition). Volume 4 : Démonstrations et résumés d'articles internationaux*, p. 22–25, Nancy, France, 2020. ATALA. URL https://hal.science/hal-02768500.

[36] G. Lecorvé, H. Ayats, B. Fournier, J. Mekki, J. Chevelu, D. Battistelli et N. Béchet : Construction conjointe d'un corpus et d'un classifieur pour les registres de langue en français. *In Traitement automatique du langage naturel (TALN)*, p. 143–156, Rennes, France, mai 2018. ATALA. URL https://inria.hal.science/hal-02002601.

[37] J. Chevelu, D. Lolive, S. Le Maguer et D. Guennec : Se concentrer sur les différences : une méthode d'évaluation subjective efficace pour la comparaison de systèmes de synthèse. *In Journées d'études sur la Parole*, Paris, France, juil. 2016. URL https://inria.hal.science/hal-01338918.

[38] J. Chevelu, G. Lecorvé et D. Lolive : ROOTS : un outil pour manipuler facilement, efficacement et avec cohérence des corpus annotés de séquences. *In Actes des Journées d'Études sur la Parole (JEP)*, Le Mans, France, juin 2014. URL https://inria.hal.science/hal-00975897.

[39] J. Chevelu, Y. Lepage, T. Moudenc et G. Putois : L'évaluation des paraphrases : pour une prise en compte de la tâche. *In Actes de Traitement Automatique des Langues Naturelles (TALN)*, Montréal, Canada, juil. 2010. ATALA. URL http://www.iro.umontreal.ca/~felipe/TALN2010/Xml/Papers/all/taln2010_submission_122.pdf.

[40] P. Alain, N. Barbot, O. Boëffard, J. Chevelu et A. Delhay : Evaluation de méthodes de réduction de corpus linguistiques. *In Actes des 27èmes Journées d'Etudes sur la Parole (JEP)*, Avignon, France, juin 2008. Association Francophone de la Communication Parlée (AFCP). URL https://hal.science/hal-04650429.

Dépots Logiciels

[41] J. Chevelu : Sc-solver : une boite à outil pour résoude les problèmes de couvertures d'ensemble par algorithme glouton, 2015. Dépot APP sous licence LGPL.

[42] D. Lolive, J. Chevelu et D. Guennec : Expression tts : moteur de synthèse par sélection d'unités acoustiques, 2015. Dépot APP.

[43] O. Boëffard, S. Lemaguer, J. Chevelu, G. Lecorvé et D. Lolive : Roots : un outil pour manipuler facilement, efficacement et avec cohérence des corpus annotés de séquences, 2014. URL http://roots-toolkit.gforge.inria.fr/. Dépot APP sous licence LGPL.

Démonstrations

[44] D. Lolive, E. Delais-Roussarie et J. Chevelu : Phorevox : apprentissage des langues par l'usage de technologies vocales. *In Session de démonstration des Journées d'Études sur la Parole (JEP)*, Le Mans, France, juin 2014. URL http://www-lium.univ-lemans.fr/jep2014/.

[45] N. Dehghani, H. Hajipoor, J. Chevelu et G. Lecorvé : Controllable paraphrase generation with multiple types of constraints. *In Demonstration in NeurIPS workshop on Controllable Generation for Modeling in Language and Vision (CtrlGen)*, 2021. URL https://ctrlgenworkshop.github.io/accepted_demos.html.