

Candidature pour la thèse intitulée « CAUSEMO » :
**Identifier automatiquement les émotions et
leurs causes dans les textes**

Gustave CORTAL

Étudiant-ingénieur à l'Université de Technologie de Compiègne (UTC)
en stage à l'ENS Paris-Saclay

Sommaire

1. Étudiant-ingénieur à l'UTC (2017 - 2022)
2. Recherches effectuées (2020 - 2022)
3. Stage de recherche au LMF+LISN (Fév. - Juil. 2022)
4. Projet : thèse en TAL au LMF+LISN

Diapositives : <https://gustavecortal.com/data/diapo.pdf>

Étudiant-ingénieur à l'UTC, Compiègne (2017 - 2022)

2017 - 2022 : Étudiant-ingénieur à l'UTC. Mathématiques, informatique et sciences cognitives. Filière science des données et intelligence artificielle. GPA : 4,64/5.

Expériences internationales

- ▶ Fev. - Juil. 2019 : Semestre d'échange à l'Université de Technologie Sino-Européenne de l'Université de Shanghai, Chine.
- ▶ Fev. - Juil. 2020 : Semestre d'échange à l'Université Nationale des Sciences et des Technologies de Taïwan, Taïwan.
- ▶ Jui. - Aou. 2022 : Oxford ML School (taux d'acceptation 5%) et Creatathon (INRIA, ENS Paris-Saclay)



Stage d'ingénieur en TAL (Makezu, Sep. - Fév. 2021)

Sujet : TAL pour le profilage des utilisateurs et l'analyse des tendances

- ▶ Makezu (Station F) : Identification des besoins et moments de vie sur les réseaux sociaux
- ▶ Développement d'un outil de partage automatique d'articles sur Twitter
- ▶ Développement d'un outil d'extraction de questions et réponses Quora
- ▶ Affinement de modèles de langage pré-entraînés (BERT) pour catégoriser les utilisateurs et les tendances

Rapport de stage : [lien](#)

Projets de recherche à l'UTC

Jan. - Juil. 2021 : Projet de recherche en sciences sociales computationnelles.

Sujet : Analyse exploratoire des réactions aux résultats d'affectation de Parcoursup sur Twitter.

Encadré par Anne Bellon et Michael Vicente (professeurs à l'UTC)

Sep. - Jan. 2022 : Projet de recherche en apprentissage automatique.

Sujet : Représentation de code HTML par l'apprentissage automatique de vecteurs.

Encadré par Sylvain Rousseau et Benjamin Quost (professeurs à l'UTC) et Sylvain Marchienne (ingénieur de recherche en apprentissage automatique chez Dashlane)

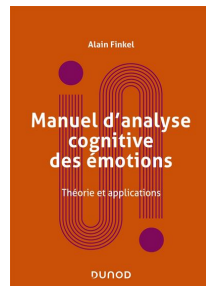
Stage de recherche au LMF+LISN (Fév. - Juil. 2022)

Sujet : Identifier les émotions dans des récits narratifs en combinant psychologie et apprentissage profond (6 mois).

Supervisé par Alain Finkel (ENS Paris-Saclay, LMF), Patrick Paroubek (LISN, CNRS) et Lina Ye (CentraleSupélec, LMF)

Stage de 5^{ème} année (éq. M2) à l'UTC

- ▶ État de l'art sur la classification jointe des émotions et de leurs causes
- ▶ Traitement des émotions basé sur les grilles FERA (Fait Émotion Raison Action)
- ▶ Outil d'extraction et de pré-traitement des grilles FERA et environnement d'annotation
- ▶ Outil de visualisation présentant différents plans d'analyse

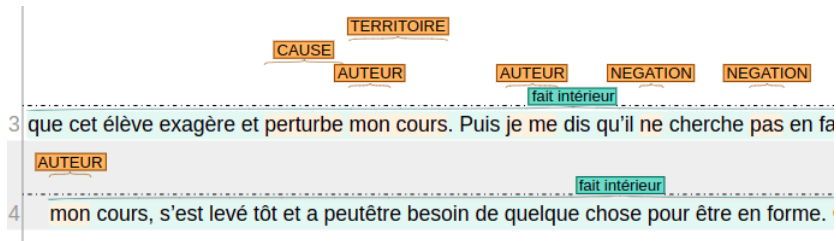


FINKEL, avril 2022

Annotation des grilles (INCEpTION)

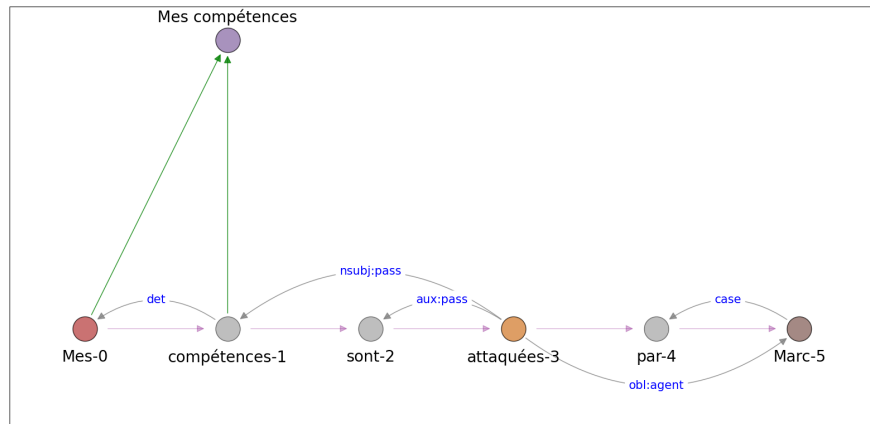
Guide d'annotation : [lien](#) Code source : [lien](#) INCEpTION : [lien](#)

1200 grilles annotées



Visualisation du graphe de relations (NetworkX)

Code source : [lien](#) NetworkX : [lien](#)



Projet : Thèse en TAL au LMF+LISN

État de l'art

Ying CHEN et al. "Emotion Cause Detection with Linguistic Constructions". In : *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. 2010

Lin GUI et al. "Event-Driven Emotion Cause Extraction with Corpus Construction". In : *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016

Rui XIA et Zixiang DING. "Emotion-Cause Pair Extraction : A New Task to Emotion Analysis in Texts". In : *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019

Penghui WEI, Jiahao ZHAO et Wenji MAO. "Effective Inter-Clause Modeling for End-to-End Emotion-Cause Pair Extraction". In : *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020

Projet : Thèse en TAL au LMF+LISN

Limites

- ▶ Jeux de données majoritairement en chinois
- ▶ Approche *pipeline* pas aussi efficace qu'une approche *end-to-end*
- ▶ Absence de communication entre la psychologie et le TAL
- ▶ Complexité importante des modèles d'extraction de n-uplets

Objectifs : Analyse jointe des émotions et des causes

- ▶ Différentes granularités
- ▶ Contexte français
- ▶ Application à l'Analyse Cognitive des Émotions (FERA)

Projet : Thèse en TAL au LMF+LISN

Planning

- ▶ 1^{er} oct. - 31 déc. 2022 : Jeu de données en français
- ▶ jan. - déc. 2023 : Étude de l'expression des émotions par des graphes
- ▶ jan. - déc. 2024 : Modèle basée sur les graphes pour identifier les émotions et leurs causes
- ▶ jan. - oct. 2025 : Modèle qui relie la perception d'une scène émotionnelle aux possibilités d'actions

Applications dans le domaine psychologique, littéraire, médical, etc.

Merci pour votre attention

Diapositives : <https://gustavecortal.com/data/diapo.pdf>

Bibliographie complétée (1)

Ying CHEN et al. "Emotion Cause Detection with Linguistic Constructions". In : *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. 2010

Lin GUI et al. "Event-Driven Emotion Cause Extraction with Corpus Construction". In : *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016

Rui XIA et Zixiang DING. "Emotion-Cause Pair Extraction : A New Task to Emotion Analysis in Texts". In : *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019

Penghui WEI, Jiahao ZHAO et Wenji MAO. "Effective Inter-Clause Modeling for End-to-End Emotion-Cause Pair Extraction". In : *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020

Aaditya SINGH et al. *An End-to-End Network for Emotion-Cause Pair Extraction*. 2021. arXiv : 2103.01544 [cs.CL]

Bibliographie complétée (2)

Muhammad ABDUL-MAGEED et Lyle UNGAR. "EmoNet : Fine-Grained Emotion Detection with Gated Recurrent Neural Networks". In : (juill. 2017), p. 718-728. DOI : [10.18653/v1/P17-1067](https://doi.org/10.18653/v1/P17-1067). URL : <https://aclanthology.org/P17-1067>

Yuan YAO et al. "DocRED : A Large-Scale Document-Level Relation Extraction Dataset". In : *CoRR* abs/1906.06127 (2019). arXiv : 1906.06127. URL : <http://arxiv.org/abs/1906.06127>

Hassan ALHUZALI et Sophia ANANIADOU. "SpanEmo : Casting Multi-label Emotion Classification as Span-prediction". In : *CoRR* abs/2101.10038 (2021). arXiv : 2101.10038. URL : <https://arxiv.org/abs/2101.10038>

Alain FINKEL. *Manuel d'analyse cognitive des émotions : Théorie et applications*. Paris : Dunod, 2022

Wenxuan ZHANG et al. "A Survey on Aspect-Based Sentiment Analysis : Tasks, Methods, and Challenges". In : (2022). DOI : [10.48550/ARXIV.2203.01054](https://doi.org/10.48550/ARXIV.2203.01054). URL : <https://arxiv.org/abs/2203.01054>

Annexe : Sciences cognitives et philosophie

Participation aux séminaires de recherche interdisciplinaires des sciences et technologies cognitives (PHITECO), chaque année de 2018 à 2022

Dissertations en philosophie des techniques sur Gilbert Simondon (2019, [lien](#)), en épistémologie sur Gaston Bachelard (2020, [lien](#)), et en cybernétique sur W. Ross Ashby (2021, [lien](#))

Annexe : Projets de recherche à l'UTC

Sujet : Analyse exploratoire des réactions aux résultats d'affectation de Parcoursup.

Développement d'outils pour la recherche de mes encadrants

- ▶ Création d'une base de données de discours.
- ▶ Élaboration de modèles de sujets.

Sujet : Représentation de code HTML par l'apprentissage automatique de vecteurs.

Représentation de code HTML par l'apprentissage automatique de vecteurs

- ▶ Étude d'une approche qui considère le code comme une structure en utilisant des réseaux de neurones en graphes.
- ▶ Étude d'une approche qui considère le code comme du simple texte en utilisant des modèles de langage pré-entraînés (GPT).

Annexe : Projets personnels sur la vision et le langage

Exploration des derniers modèles de vision par ordinateur pour produire de l'art depuis 2020 (gustavecortal.com)

Affinement de modèles de langage pré-entraînés sur mes jeux de données personnels pour produire du texte philosophique (2021, [lien](#)) et des articles français (2022, [lien](#))

Quantification 8-bits de modèles de langage pré-entraînés (400 téléchargements par mois) (2022, [lien](#))

Publication d'un jeu de données français contenant 40.000 articles de 50 journaux francophones parlant de la COVID-19 (2022, [lien](#))