Criminologie, Forensique et Sécurité

www.revue-cfs.net | ISSN : 2813-6098

Organe officiel de l'association internationale des criminologues de langue française (AICLF)

Article régulier

Criminologie, Forensique et Sécurité

2026 | 4 (1) : 8332

https://doi.org/10.26034/la.cfs.2026.8332

L'analyse des traces langagières comme vecteur de détection de relations entre des cas de fraudes en ligne

Clara Degeneve[1], **Julien Longhi**[2] et **Quentin Rossy**[3]

[1, 3] École des Sciences Criminelles, University of Lausanne, 1015 Lausanne, Suisse
[2] AGORA Laboratory EA 7392, CY Cergy Paris University, 95000 Cergy, France

Contact : clara.degeneve@unil.ch

Résumé

Cette étude examine dans quelle mesure l'analyse des traces langagières, entendues ici comme des éléments textuels présents dans des textes illégaux ou litigieux, pouvant renseigner non seulement sur leur source, mais également sur l'activité d'intérêt elle-même, permet d'identifier des relations entre des cas de fraudes amoureuses et d'éclairer le fonctionnement du système de délinquance qui les structure. En raison de l'homogénéité du corpus analysé et du caractère fortement scripté des interactions frauduleuses, l'analyse des segments répétés a été retenue afin de mettre en évidence des similitudes présentant un pouvoir discriminant suffisant. Les résultats révèlent l'usage récurrent de « phrases modèles » reproduites dans plusieurs cas, indiquant la circulation de scripts standardisés au sein de communautés de fraudeurs. Des recherches en ligne fondées sur des segments caractéristiques ont permis d'identifier des plateformes dédiées au partage de ces scripts, suggérant l'existence d'espaces virtuels de convergence favorisant la diffusion et l'appropriation de modèles. L'analyse des traces langagières ne se limite ainsi pas à relier certains cas entre eux : elle offre également un accès privilégié aux dynamiques d'interaction, de concentration et de circulation des informations dans le système de délinquance associé aux fraudes amoureuses.

Mots clés

Trace langagière, fraude amoureuse, fraude en ligne, espace virtuel de convergence et analyse criminelle

The analysis of linguistic traces as a means of detecting links between online fraud cases

Abstract

This study explores the extent to which the analysis of language traces, understood here as textual elements in illegal or disputed texts that may provide information not only about their source but also about the activity of interest itself, can be employed to identify relationships between romance fraud cases and to shed light on the structure and functioning of the criminal system underlying these offences. Due to the homogeneity of the corpus analysed and the highly scripted nature of fraudulent interactions, repeated-segment analysis was selected to highlight similarities with sufficient discriminatory value. The findings reveal the recurrent use of "template phrases" reproduced across multiple cases, indicating the circulation of standardised scripts within fraudster communities. Searches conducted online using characteristic segments further enabled the identification of platforms dedicated to the sharing of such scripts, suggesting the existence of virtual convergence spaces that support the dissemination and uptake of templates. Language trace analysis therefore not only enables the connection of individual cases but also provides insight into the dynamics of interaction, and circulation of information that characterise the behaviour system associated with romance fraud.

Keywords

Linguistic trace, romance fraud, online fraud, virtual convergence space and crime analysis

Citation : Degeneve, C., Longhi, J., et Rossy, Q. (2026). L'analyse des traces langagières comme vecteur de détection de relations entre des cas de fraudes en ligne. *Criminologie, Forensique Et Sécurité*, *4*(1) : 8332.

Introduction

À mesure que les interactions sociales, économiques et commerciales se déplacent vers l'espace numérique, la fraude s'y reconfigure. Ce déplacement, plus que technologique, transforme les vulnérabilités des lésés, les modalités d'action des auteurs et les possibilités d'enquête (Reep-van Den Bergh et Junger, 2018). Dans ce contexte, la pandémie de COVID 19 a constitué un accélérateur de fraudes en ligne (Buil-Gil et Zeng, 2022). Entendue au sens large, la fraude peut se définir comme « un acte de tromperie visant à priver les victimes de leur argent, de leurs données ou d'autres biens. » (Rossy et Ribaux, 2020, p.336).

Parmi la diversité de formes, les fraudes amoureuses, également appelées « romance scams », exploitent la vulnérabilité émotionnelle et la solitude des individus pour des gains financiers (Whitty, 2015; Whitty et Buchanan, 2012). Ces escroqueries, notamment orchestrées par des groupes criminels basés en Afrique de l'Ouest (Atta-Asamoah, 2009), reposent sur des interactions en ligne intégrant des stratégies destinées à établir la confiance et manipuler les victimes jusqu'à les amener à verser ou blanchir de l'argent sur demande du criminel (Carpentier *et al*., 2024; Rossy et Ribaux, 2020; Whitty, 2015). En Suisse, la police enregistre plusieurs centaines de cas chaque année (Akermann, 2025).

La prise de contact entre le fraudeur et le (la) lésé.e se fait, le plus souvent, sur des plateformes de rencontre. Après une première phase d'échanges écrits, les fraudeurs cherchent généralement à déplacer la conversation sur des services de messagerie privés, tels que WhatsApp, Telegram, ou les emails afin de s'extraire de la surveillance des plateformes initiales (Carter, 2024; Wang et Topalli, 2024; Whitty, 2015). Les échanges se multiplient ensuite et peuvent s'étendre sur la durée (Whitty et Buchanan, 2012), laissant une importante quantité de traces des conversations pouvant servir de données de base pour la reconstruction des modes opératoires.

Si la plupart des analyses linguistiques menées sur les fraudes amoureuses portent sur les traces produites lors de ces échanges, elles visent principalement à comprendre les mécanismes de persuasion mobilisés par les fraudeurs ainsi que les facteurs de victimisation (Buchanan et Whitty, 2014; Kopp *et al*., 2016; Lazarus *et al*., 2023). D'autres travaux se concentrent sur la détection de faux profils à partir des informations accessibles sur les sites de rencontre (De Jong, 2019; Toma et Hancock, 2012). En revanche, aucune étude adoptant une approche linguistique ne s'est, à notre connaissance, intéressée à la détection de relation entre ces cas de fraude. Ainsi, cette recherche propose d'explorer le potentiel d'exploitation des méthodes issues de la linguistique pour comprendre les systèmes de délinquances liés aux fraudes amoureuses. Un système de délinquance peut se définir comme « un ensemble coordonné de pratiques et d'interactions grâce auxquelles des malfaiteurs exploitent les occasions offertes par une brèche dans les protections contre le crime ou créent une telle brèche. » (Cusson, 2005 citantTremblay, 2004, p.94).

Dans ce contexte, l'analyse des traces langagières (Degeneve *et al*., 2022 ; Renaut *et al*., 2017) émerge comme un outil pour détecter des relations entre les cas et inférer l'activité de groupes d'individus. La recherche se focalise sur deux questions principales :

> Est-il possible de détecter des relations entre des cas de fraude amoureuse en se basant uniquement sur l'analyse de textes écrits aux victimes par les fraudeurs ?
>
> Dans quelle mesure l'analyse des traces langagières permet-elle la détection d'espaces d'échange en ligne entre les fraudeurs ?

La recherche s'appuie sur des conversations entre des fraudeurs et des chercheur.eus.es de l'Institut de Lutte contre la Criminalité Economique de Neuchâtel (Zbinden *et al*., 2023). Elle vise, dans un premier temps, à mettre en évidence les traces textuelles pertinentes pour détecter des relations entre les cas. L'objectif n'est pas ici d'établir une distinction entre profil légitime et profil frauduleux. Il s'agit ici de détecter des relations entre les cas afin d'avoir une meilleure compréhension du système de délinquance. À cette fin, une analyse des segments répétés est menée sur l'ensemble des messages envoyés par les fraudeurs. Chaque conversation est considérée comme un cas distinct. Si des segments issus du langage courant, comme des formules de politesse, peuvent être considérés comme peu discriminants, des segments longs et complexes peuvent par hypothèse s'avérer plus spécifiques et pertinents. Dans un second temps, l'hypothèse est posée que les segments pertinents détectés peuvent être exploités comme critère de recherche pour détecter des espaces d'échanges en ligne, appelés des espaces de convergence en ligne, (Soudijn et Zegers, 2012) où les fraudeurs échangent des scripts de communication (Asyalı *et al*., 2026; Faber, 2024) et des exemples de discussion. Ces espaces se présentent majoritairement sous la forme de blogs ou de forums de discussion librement accessibles sur le web et contiennent des modèles de conversation.

Cette recherche s'inscrit ainsi à l'intersection de la science forensique, de la linguistique et de la criminologie, en mobilisant des approches complémentaires pour analyser la fraude en ligne et en élargir les perspectives d'exploitation. Les traces textuelles issues de l'activité des fraudeurs sont examinées par des approches computationnelles de la linguistique, afin de mettre en évidence des similitudes langagières desquelles des relations sont inférées pour reconstruire les activités sérielles. La pertinence des segments répétés identifiés est évaluée en regard de la nature de l'activité criminelle, appréhendée selon une approche par script issue des études criminologiques sur le phénomène. De plus, la reconnaissance de segments de discussion caractéristiques de la fraude vise la détection des espaces de convergence mobilisés par les fraudeurs. La démarche permet ainsi de mettre à jour d'autres traces utiles à la reconstruction des dynamiques d'échange et des pratiques délinquantes. Par hypothèse, l'identification de scénarios récurrents et des mécanismes de partage peuvent alors constituer un socle pour l'élaboration de stratégies de sensibilisation et de perturbation.

Revue de la littérature

La fraude amoureuse

Déroulement de la fraude amoureuse

La fraude amoureuse est un type particulier de fraude en ligne consistant à établir et entretenir une relation amoureuse avec une victime dans le but de lui soutirer de l'argent (Whitty, 2019; Whitty et Buchanan, 2012). Dans l'analyse proposée par Faber (2024), ce processus peut être appréhendé comme une

transaction commerciale frauduleuse, dans laquelle le fraudeur «vend» une relation amoureuse à la victime, qui en paie progressivement le maintien.

Des recherches récentes (Carter, 2024; Faber, 2024; Schokkenbroek et Snaphaan, 2025) proposent un script de déroulement de la fraude amoureuse en se basant sur le travail de Whitty (2015), dont les étapes principales peuvent être résumées comme suit :

- *Établissement du faux profil et prise de contact* : le fraudeur crée un faux profil attractif et entre en contact avec la victime à travers un site de rencontre ou un réseau social.
- *Développement de la relation et test de confiance* (« *Grooming* ») : le fraudeur développe la relation avec la victime sur un laps de temps variable, déclarant son amour et laissant entrevoir la perspective d'une relation sur le long terme. C'est généralement à cette étape que le fraudeur propose un déplacement de la conversation du site de rencontre vers des applications de messagerie instantanée ou des mails. Si des rencontres physiques sont proposées, elles n'auront jamais lieu. Faber (2024) affine l'étape en y distinguant plusieurs sous-étapes : l'introduction, les informations personnelles et intérêts partagés, la mise en place du registre amoureux, la déclaration d'amour, puis sa confirmation et sa consolidation.
- *Crise* (« *The Sting* ») *et remise de bien* : lorsque le fraudeur considère la victime prête, il tente de lui soutirer de l'argent sous divers prétextes. Plusieurs schémas d'extorsion sont possibles, allant des petites sommes ponctuelles à une somme très importante d'un coup. Le fraudeur peut également demander à la victime de transférer de l'argent d'un compte à un autre ou de stocker de l'argent sur son propre compte, l'impliquant dans des cas de blanchiment d'argent (Huhn, 2023).
- *Chantage* (« *Sexual abuse* ») : cette étape est plus rare, mais peut impliquer un chantage si des images ou vidéos compromettantes ont été transmises au fraudeur par la victime lors de leurs échanges.
- *Révélation* : la victime réalise la fraude, soit d'elle-même, soit en étant informée par les autorités. Le signalement à la police n'est pas systématique, la victime pouvant se sentir honteuse d'avoir été trompée ou redoutant d'avoir été impliquée dans des activités illégales, comme le blanchiment d'argent.

Ainsi, chaque étape pourrait présenter ses propres caractéristiques linguistiques en fonction de l'objectif poursuivi par le fraudeur, suggérant un type de discours différent pour chaque étape de l'arnaque (Faber, 2024). Par exemple, selon Kopp *et al*. (2016), la création d'un faux profil se fait selon des critères connus pour attirer les victimes. La construction d'une histoire crédible constitue ensuite une étape cruciale afin de pouvoir mener le script de la fraude jusqu'à l'extorsion d'argent. La victime doit se sentir impliquée dans cette histoire, le fraudeur lui faisant croire qu'elle correspond à son « idéal » amoureux et adaptant son discours en fonction de son profil : « la fraude amoureuse suit le scénario de l'histoire d'amour personnelle de la victime » (Kopp *et al*., 2016 :215).

Victimes, auteurs et espaces de convergence

Le préjudice de la fraude amoureuse pour les victimes est double, puisqu'en plus de subir une perte d'argent souvent importante, elles subissent une perte émotionnelle liée à la relation construite avec le fraudeur (Carpentier *et al*., 2024; Whitty et Buchanan, 2012). Le sentiment de honte peut empêcher le report à la police, ce qui ne permet pas d'avoir une estimation exhaustive du nombre de cas réels par les données policières (Akermann, 2025; Anesa, 2020; Carpentier *et al*., 2024).

La revue de littérature effectuée par Lazarus *et al*. (2023) met en évidence que la plupart des recherches sur les fraudes amoureuses sont des études socio-criminologiques centrées soit sur la victime, soit sur l'auteur. Les études portent sur les caractéristiques de chacun pour comprendre le phénomène de victimisation et son impact, ainsi que les motivations des auteurs et leurs techniques de neutralisation. Dans le cadre de cette recherche, les aspects liés aux victimes ne seront pas abordés.

Les fraudeurs à l'origine des fraudes amoureuses semblent pour la plupart localisés dans des pays d'Afrique de l'Ouest (Barnor *et al*., 2020), ou en Chine, dans des structures plus organisées où la fraude est appelée « Pig Butchering » et intègre des mécanismes de fraudes à l'investissement (Asyalı *et al*., 2026). Alors que les individus avaient initialement tendance à travailler de manière autonome, on observe aujourd'hui une augmentation des groupes organisés, au sein desquels chaque membre occupe une place définie (Atta-Asamoah, 2009). Les fraudeurs qui travaillent en groupe exploitent des compétences propres à chacun (Cretu-Adatte, Azi, *et al*., 2024; Soares et Lazarus, 2024). Ils peuvent également s'appuyer sur des manuels fournissant des instructions et conseils pour mener la fraude à son terme (Asyalı *et al*., 2026).

Les compétences techniques sont principalement acquises par l'apprentissage entre pairs, ce qui suppose des échanges soutenus entre les individus. Les auteurs de fraudes amoureuses basés en Côte d'Ivoire semblent se retrouver dans un premier temps dans le monde physique (Cybercafés, habitations, parfois prison) et travaillent généralement avec des connaissances, des amis ou des membres de leur famille (Cretu-Adatte, Azi, *et al*., 2024). Un vocabulaire spécifique semble partagé par ces auteurs, qui utilisent des termes du nouchi ivoirien, tels que « mougou », qui désigne le « client »,c'est-à-dire la personne à piéger (Adou, 2022). L'usage du terme « client » (customer) dans les manuels de pig butchering étudiés par Asyalı *et al*. (2026) témoigne d'une conceptualisation marchande de la victime, en cohérence avec l'analyse de Faber (2024), qui appréhende la fraude amoureuse à travers le prisme d'une transaction commerciale.

Les échanges entre fraudeurs sont également facilités par des espaces de convergence en ligne, comme les forums décrits par Soudijn et Zegers (2012). Il s'agit de plateformes qui permettent aux criminels de partager des connaissances, des informations et de proposer des produits ou des services. Le recrutement d'intermédiaires ou de mules impliqués dans

le blanchiment d'argent ou la récupération de biens peut en outre s'opérer par ces canaux, mais également via des applications de messagerie instantanée ou dans le cadre de rencontres en présentiel. D'autres formes de rencontres visent enfin à corrompre des agents administratifs (employés de banque, policiers, etc.) afin de faciliter les activités illicites.

La prévention et la perturbation des fraudes amoureuses

Les mesures de prévention se basent principalement sur des campagnes d'information qui doivent être suffisamment générales pour que les utilisateurs puissent identifier les fraudes dans leur propre situation (Kydd *et al*., 2024).

Bilz *et al*. (2023) proposent deux approches distinctes :

- *Approches machine* : détection automatique des profils frauduleux par comparaison automatique d'images et de textes du profil public. Les analyses de base ne semblent pas montrer une efficacité satisfaisante (De Jong, 2019; Suarez-Tangil *et al*., 2019). Le problème majeur dans ces recherches reste l'accessibilité à un jeu de données. La plupart des études reposent sur le même jeu de données, introduisant potentiellement des biais lors des analyses. Les méthodes par apprentissage machine sont également difficilement exploitables par les utilisateurs des plateformes en ligne. Elles doivent être mises en place par les gestionnaires des plateformes.
- *Approches humaines* : afficher des messages de mise en garde pour indiquer des consignes aux utilisateurs, comme le fait de se méfier lorsqu'une demande d'argent intervient dans la conversation. L'approche humaine de prévention nécessite une dimension proactive et consciente de la part des victimes potentielles. À travers 52 témoignages de victimes de fraude amoureuse postés sur des sites de signalement de fraude, Wang & Topalli (2024) ont mis en évidence le fait que, malgré des signes inquiétants de la part de leur interlocuteur, certaines victimes ont été piégées. Celles-ci racontent avoir eu conscience que quelque chose n'était pas normal, mais avoir accédé aux demandes du fraudeur malgré leurs doutes par peur de perdre leur relation. Suite à cette expérience, les victimes n'ayant pas subi de perte financière seraient moins enclines à signaler le cas à la police. Afin d'aider les victimes à se reconstruire et en prévenir de nouvelles fraudes, il existe des groupes de soutien dans lesquelles celles-ci peuvent partager leur expérience (Wang et Topalli, 2024). Les sites de signalement contribuent ainsi à identifier des indices de fraude, grâce au partage d'expériences des utilisateurs.

En complément des approches préventives, le « scambaiting », que nous appellerons du « leurrage d'escrocs », peut également être envisagé comme une approche de perturbation (Smart, 2025; Sorell, 2019; Zbinden *et al*., 2023). Leurrer un escroc peut se décrire comme une pratique de vigilantisme numérique consistant à échanger des messages avec des auteurs de fraude, généralement dans le but de leur faire perdre du temps ou de récupérer des informations utiles à la reconstruction des activités sérielles et de détection des cas (Zingerle et Kronman, 2018). En effet, Sorell (2019) met en évidence que l'action des piégeurs d'escroc civils permet non seulement de perturber les criminels en leur faisant perdre du temps, mais également de récolter des informations qu'ils peuvent partager ensuite via des sites spécialisés dans la dénonciation ou sur les réseaux sociaux afin d'aider les victimes à reconnaître la fraude. Par exemple, lors de leur étude menée entre 2021 et 2022, Zbinden *et al*. (2023) ont collecté des données à partir de conversations avec des fraudeurs impliqués dans des fraudes amoureuses en procédant à du leurrage d'escrocs. Des informations identifiantes ont été extraites des traces d'échanges, comme des numéros de téléphone, des adresses mail, des adresses IP, des comptes bancaires et des adresses Bitcoin. Pris isolément ou combinés, ces éléments peuvent permettre d'individualiser une entité virtuelle à un moment donné, par exemple le correspondant d'une communication téléphonique ou le compte d'arrivée d'une transaction financière. Cette entité peut renvoyer indirectement à une personne physique ou à un dispositif technique. Dans les textes légaux, la notion de « donnée personnelle » vise les informations identifiantes relatives à une personne physique. Le concept d'information identifiante est toutefois plus large, puisqu'il peut aussi s'appliquer à des entités virtuelles ou à des identités fausses ou usurpées (Bollé, 2025: 44). La plupart de ces informations sont transmises par le fraudeur lui-même dans le cadre d'une conversation, à l'exception des adresses IP par exemple, qui sont extraites des traces observées. Dans le cadre de l'étude, ces dernières ont été récoltées lorsque le fraudeur se connectait à un faux site internet de location transmis par le chercheur.

L'exploitation des traces langagières pour reconstruire la fraude amoureuse

La fraude amoureuse, comme la plupart des fraudes en ligne, est susceptible de laisser des traces numériques détectables dans le cadre d'une investigation. Une trace peut se définir comme « toute modification de l'environnement, observable ultérieurement, résultant d'un événement » (Pollitt *et al*., 2018, p.1, traduction libre de l'anglais).

Dans l'environnement numérique, le texte écrit demeure l'un des principaux vecteurs de communication, que ce soit à travers des publications sur les réseaux sociaux, des conversations sur des applications de messagerie ou encore des annonces sur des plateformes de vente. Ces éléments textuels pertinents pour l'analyse sont nommés ici traces langagières. Celles-ci peuvent se définir comme l'écriture d'un texte illégal ou litigieux par un auteur et qui présente un potentiel informatif non seulement sur sa source, mais également sur l'activité illicite elle-même (Degeneve *et al*., 2022).

Les usages de ces traces langagières recensés dans la littérature sont multiples. Elles peuvent être, par exemple, exploitées afin de détecter des cas de désinformation (Addawood *et al*., 2019; Rubin, 2016) ou la publication d'annonces frauduleuses sur les marchés en ligne (Degeneve *et al*., 2024). Zhou *et al*. (2004) proposent une approche stylo-métrique destinée à détecter les tromperies lors d'échanges par mail entre deux interlocuteurs. Des caractéristiques comme le nombre de mots, de phrases, l'expressivité, la diversité lexicale, les fautes typographiques ou encore l'expression de l'incertitude ont été comparées statistiquement pour déterminer quels sujets de l'étude appartenaient au groupe des menteurs. Dans une veine similaire, Addawood *et al*. (2019) ont mis en évidence un lexique d'indicateurs linguistiques pouvant être considérés comme marqueurs de textes conçus pour induire en erreur les lecteurs, dans le cadre d'une étude sur les soupçons de

manipulation de l'élection présidentielle américaine de 2016.

Bien que les recherches portant sur l'exploitation des traces langagières se soient largement concentrées sur la différenciation entre informations légitimes et désinformation, notamment sur les réseaux sociaux et dans la presse, les cas de fraude amoureuse peuvent eux aussi être abordés sous l'angle de l'analyse linguistique. Dans le cas des escroqueries en ligne, les auteurs peuvent échanger par écrit avec les victimes durant des mois, voire des années, laissant de nombreuses traces langagières utiles à l'investigation. Toma et Hancock (2012) proposent une analyse des profils d'utilisateurs de sites de rencontre, visant à détecter les menteurs à partir du contenu de ces profils. En effet, comme exprimé précédemment, les escrocs construisent un faux profil de toute pièce. Chaque participant de l'étude est invité à évaluer la véracité des éléments descriptifs des profils, comme la taille, le poids, l'âge et la photographie. Les profils écrits sont ensuite analysés à l'aide de la méthode « Linguistic inquiry and word count » (LIWC) (Tausczik et Pennebaker, 2010), qui repose sur une comparaison de mots à partir d'un dictionnaire préconstruit. La recherche a mis en évidence une surutilisation d'indicateurs émotionnels positifs dans les profils de menteurs, mais également des descriptions moins longues et moins complexes ainsi qu'une esquive des sujets ayant trait aux éléments descriptifs lorsque les utilisateurs n'ont pas été honnêtes sur l'élément en question.

L'analyse des traces langagières permet également d'informer sur la structure d'un texte, de détecter des récurrences ou de caractériser une forme de discours. Le schéma lexico-grammatical correspondant à une unité du discours de taille variable dont la récurrence est observable vise alors à caractériser la structure d'un discours (Gledhill *et al*., 2017; Halliday, 2014). Ils permettent notamment la mise en évidence deux types principaux de structure: la macro-structure, une structure grammaticale stable et observable de manière récurrente, la micro-structure, comportant des éléments plus variables selon les différents textes dans le corpus (Gledhill *et al*., 2017).

Des analyses de discours ont également été menées sur les méthodes de persuasion employées par les fraudeurs pour piéger leurs victimes dans les fraudes amoureuses (Carter, 2024; Lee *et al*., 2023). Dans leur recherche, Lee *et al*. (2023) proposent une approche comparative entre des profils frauduleux (N = 500) obtenus à partir d'une plateforme de signalement et des profils réels (N = 500) issus de cinq sites de rencontre distincts. Une analyse du vocabulaire le plus fréquent pour chacune des deux catégories a permis de mettre en évidence un champ lexical plus fréquent chez les profils de fraudeurs : la description de soi et du.de la partenaire idéal.e, à l'aide d'adjectifs positifs. Les entités et noms détectés plus fréquemment dans les profils de fraudeurs sont consistants avec la thématique de description de soi et du/de la partenaire décrite précédemment. Le terme « God » est, par exemple, significatif dans la distinction des fraudeurs par rapport aux profils légitimes (Anesa, 2020; Carter, 2024; Faber, 2024; Lee *et al*., 2023). Les termes liés à la religion sont, selon les auteurs, employés par les fraudeurs afin de renforcer leur image de personne dévouée, généreuse et moralement droite. Ils vont parfois jusqu'à attribuer leur rencontre avec la victime à un signe de Dieu, contribuant à inspirer la confiance chez elle.

De nombreux auteurs ont mis en évidence l'usage de « scripts de conversation » suivis par les fraudeurs lors de leurs échanges avec les victimes (Anesa, 2020; Faber, 2024; Smart, 2025), faisant écho à la notion de structure lexico-grammaticale de (Gledhill *et al*., 2017). Ces scripts sont employés en adéquation avec les grandes étapes du déroulement de la fraude. Carter (2024) a notamment, par le biais d'une analyse de discours appliquée à des conversations entre fraudeurs et victimes, mis en évidence les stratégies employées par les fraudeurs pour entraîner les victimes vers les demandes d'argent. Chaque étape de la fraude présenterait ainsi des types de discours différents, en fonction de ce que le fraudeur veut faire ressentir à la victime à ce moment de la fraude. Lors de l'étape de développement de la relation, il va, par exemple, mettre en avant l'exception que représente la victime dans sa vie et la convaincre qu'ils ont un avenir ensemble.

Faber (2024) propose une analyse de conversations entre victimes et fraudeurs afin de mettre en évidence l'usage de ces scripts de conversation par les auteurs en se basant sur les répétitions dans leurs messages. L'auteure dispose pour cela d'un corpus de 53 échanges entre elle-même et les fraudeurs sur Facebook Messenger. Les conversations étaient poursuivies jusqu'à ce que le fraudeur demande de l'argent, et, à ce moment, l'auteure les informait qu'ils avaient été trompés et que leurs conversations seraient exploitées pour la recherche. Des modèles de conversation employés à travers plusieurs cas ont été mis en évidence à chacune des étapes de la fraude décrites par Whitty (2015). L'auteure n'émet cependant aucune hypothèse sur les potentielles relations entre les différents cas du corpus, malgré des similitudes dans les phrases employées par les fraudeurs entre différents cas.

La recherche d'Anesa (2020) porte sur l'exemple d'un fraudeur ayant perpétré des fraudes amoureuses interpellé aux États-Unis. L'analyse a permis la mise en évidence de l'usage de modèles de conversation à envoyer en fonction des réponses de la victime. La structure de ces modèles suggère qu'il s'agit de consignes réutilisables. L'analyse de discours à partir de données issues d'un fil de discussion Reddit dédié au leurrage d'escrocs a également permis à Smart (2025) de mettre en évidence la dépendance des fraudeurs à des scripts de conversation. En effet, ceux-ci ont montré, dans le déroulé des conversations analysées, peu de flexibilité face au manque de coopération des piégeurs d'escrocs refusant de suivre leurs scripts.

Si les analyses linguistiques sur les fraudes amoureuses sont majoritairement concentrées sur les mécanismes linguistiques visant à susciter la confiance et la tromperie chez la victime, aucune étude n'a encore été menée sur l'exploitation des traces langagières dans la mise en évidence de relations entre ces cas de fraude. Par ailleurs, l'accès limité à des données directement produites par les fraudeurs complique les analyses centrées sur ces derniers plutôt que sur les victimes (Abubakari, 2024; Asyalı *et al*., 2026; Smart, 2025). Ces travaux soulignent que la plupart des corpus disponibles proviennent de témoignages de victimes ou de rapports policiers, ce qui restreint l'observation du point de vue de l'auteur de la fraude. À cet égard, les données issues du leurrage d'escrocs offrent un accès privilégié à cette perspective.

Méthodologie

Hypothèses de recherche

La présente recherche a pour objectif de répondre à la problématique de la détection de relations entre des cas de fraudes amoureuses en ligne, ainsi que de renseigner sur les méthodes d'échange de connaissances employées par les

fraudeurs, contribuant au fonctionnement du système de délinquance lié à ces fraudes. Les hypothèses de travail suivantes ont été formulées :

H1: L'analyse des segments de texte issus des traces de communication des fraudeurs permet de détecter des répétitions entre des cas de fraude amoureuse.
H2 : La mise en évidence de ces segments répétés permet la détection d'espaces d'échange de modèles de conversation entre les fraudeurs disponibles sur le web.

Données

Les données exploitées dans le cadre de ce projet ont été mises à disposition par l'Institut de Lutte contre la Criminalité économique (ILCE) de Neuchâtel, en Suisse. Il s'agit de 180 cas de fraudes amoureuses, récoltés à la suite d'une étude menée sur trois mois, dans laquelle du leurrage d'escrocs a été réalisé afin de récolter des données de conversation avec des fraudeurs présumés (Zbinden *et al*., 2023). Les données à disposition sont des conversations textuelles entre les fraudeurs et les chercheurs. Les échanges sont formés de messages allant d'un mot à plusieurs phrases longues. Les données sont réparties sur trois pays, en fonction de la localisation indiquée par le.la chercheur.euse sur le site de rencontre (voir Tableau 1).

Tableau 1
Répartition des cas en fonction de la localisation de la fausse victime indiquée par les chercheur.euses de l'ILCE dans leurs profils sur les sites de rencontre

Localisation	Suisse	Canada	France	Total
Nombre de cas	54	54	72	180
Intervalle de temps	01.02.32 – 28.02.32	02.05.22 – 18.08.22	02.03.22 – 31-03-22	

La majorité des cas s'étend sur deux types de plateformes. Dans un premier temps sur un site de rencontre (Lovoo et Tinder), puis via une ou plusieurs applications de messagerie instantanée: Hangout, Skype, What's App ou Telegram. La répartition des canaux est présentée dans le Tableau 2.

Tableau 2
Répartition des canaux de communication en fonction de la localisation du profil de la victime simulée.

Canal primaire	Canal secondaire	Suisse	Canada	France
Tinder	What's App	16	0	27
	Skype	0	0	1
	Hangout	2	2	2
	Courriel	0	0	0
	Telegram	1	0	0
	Signal	0	1	0
	Viber	0	0	0
	What's App x Hangout	3	0	2
	What's App x Telegram	1	0	0
	Skype x Hangout	0	0	1
	Pas de suite	1	0	3
Lovoo	What's App	15	20	17
	Skype	0	2	1
	Hangout	8	27	12
	Courriel	0	0	0
	Telegram	0	0	0
	Signal	0	0	0
	Viber	0	2	0
	What's App x Hangout	4	0	4
	What's App x Telegram	0	0	0
	Pas de suite	3	0	2
	TOTAL	54	54	72

Prétraitements

Dans l'objectif de détecter des liens entre les cas, les messages des chercheus.e.s prenant le rôle de victime ont été filtrés afin de ne garder que les textes écrits par les fraudeurs. Les messages système liés aux applications ont également été supprimés.
L'ensemble des informations identifiantes présentes dans les conversations a ensuite été anonymisé. En plus de garantir l'éthique de la recherche par la suppression des données personnelles, cette étape d'anonymisation permet de garantir que la détection de liens entre les différents cas se base uniquement sur les données textuelles et non sur des informations identifiantes lors de l'étape d'analyse linguistique. En raison des difficultés de détection automatisée des entités nommées, tous les textes ont été lus et anonymisés manuellement selon la systématique de remplacement suivante :

- Patronyme : [[NOM]]
- Numéro de téléphone : [[TEL]]
- Adresse mail : [[MAIL]]
- Compte bancaire : [[COMPTE]]
- Pour Hangout et WhatsApp CA : [[MAIL_F]] et [[MAIL_ILCE]] ; [[NOM_F]] et [[NOM_ILCE]]
- Identifiant : [[ID]]
- Lien réseaux sociaux : [[LIEN]]
- Adresse précise : [[ADRESSE]]

Le format des conversations à disposition présentait des variations selon le canal dont elles étaient extraites, ainsi l'ensemble des traces textuelles ont été copiées au format brut dans des fichiers texte.

Détection de liens par les informations identifiantes

À la suite de l'étape d'anonymisation, une première analyse a été effectuée afin de reconstruire les liens fondés sur les informations identifiantes détectées dans les échanges.
Chaque information anonymisée dans le texte a été stockée dans une base de données selon son type (mail, numéro de téléphone, patronyme, donnée financière, etc.). Toutes ces informations ont ensuite été comparées les unes aux autres manuellement, afin de mettre en évidence des informations similaires entre différents cas et ainsi établir un premier lien

entre ces cas. À noter que les informations de profils issues de captures d'écran du site de rencontre ont également été exploitées pour la détection des liens. Dans ce cas, un lien est inféré si le patronyme, l'âge et la localisation sont concordants entre les cas. Dans le cas des patronymes, le lien est considéré comme confirmé lorsque le prénom et le nom de famille indiqués sont concordants. Le prénom ou le nom pris individuellement ne sont pas considérés comme suffisamment discriminants. Ces liens sur les informations identifiantes seront par la suite exploités afin de déterminer s'ils sont confirmés ou non par les liens détectés à l'aide des traces langagières. Ils constituent ainsi un référentiel de comparaison, une forme construite de « *ground truth* », ou « vérité terrain » reposant sur des similitudes indépendantes du contenu linguistique analysé pour inférer les relations.

Détection de liens par l'analyse des segments répétés

Il s'agit ici de la deuxième étape de détection de liens entre les cas, en se basant cette fois uniquement sur les textes anonymisés des fraudeurs.
Compte tenu de la trop faible longueur des textes à disposition dans les canaux primaires (les sites de rencontre), seuls les canaux secondaires, c'est-à-dire les applications de messagerie instantanée, ont été pris en compte dans la suite des analyses. En effet, comme exprimé dans la revue de littérature, les fraudeurs demandent très rapidement à quitter l'application de rencontre après la prise de contact.
La recherche de répétitions entre les textes est effectuée ici grâce à la méthode des segments répétés, une approche lexicométrique proposée par Salem (1986). Un segment correspond à « toutes les suites de formes graphiques non séparées par une ponctuation forte » (Salem, 1986, p.8). Il s'agit d'une forme de schéma lexico-grammatical tel que décrit par Gledhill *et al*. (2017). L'approche repose sur une comparaison de segments de taille variable afin de détecter les répétitions entre les textes. Cette méthode peut sembler analogue à une analyse des N-Grammes (Lam *et al*., 2021), très courante dans le domaine de l'analyse linguistique, mais présente l'avantage de ne pas limiter la détection à des groupes de formes (mots, caractères) de taille fixe. Le choix d'employer l'analyse des segments répétés plutôt que des méthodes statistiques plus classiques issues de la linguistique computationnelle vient majoritairement du fait que les portions de textes sont parfois très courtes, limitant grandement l'utilité d'un calcul de statistiques textuelles. De plus, ces méthodes ne permettent pas de mettre en évidence de longues séquences de mots communes entre les cas, utiles pour établir des liens.
Afin de ne pas modifier les traces, aucune approche de lemmatisation[1] n'est appliquée. D'après Bonnafous (1988), résumant la pensée de Salem (1986) « *les données issues du repérage des segments répétés permettent de lever certaines ambiguïtés entraînées par la non-lemmatisation du texte de départ* » (p. 168). C'est ainsi que l'observation de segments comme « la politique », « cette politique » ou « démocratie politique » permet de distinguer les occurrences correspondant à un emploi adjectival de celles correspondant à un emploi substantif (p. 167). Comparée à l'exploitation des formes uniques[2] ou des N-grammes, l'analyse des segments répétés permet également de prendre en compte le contexte entourant chaque forme et chaque groupe de formes. Ainsi, il a été décidé de découper les textes par itération de message des fraudeurs, chaque message envoyé représentant un segment à comparer.
Afin de procéder à cette analyse, le module Python Textacy[3] a été utilisé. Tous les segments de tous les cas ont été comparés. Seuls les segments apparaissant dans au moins deux cas ont été extraits.
Les segments constitués uniquement de signes de ponctuation, d'emojis, ainsi que les formules de politesse ou expressions usuelles (par exemple : « bonjour », « ça va », « tu fais quoi ») ont été exclus par la suite de l'analyse, car jugés trop faiblement discriminants.

Évaluation de la pertinence des liens par la catégorisation des segments

Afin de distinguer les liens pertinents et non pertinents, une classification des segments communs selon le type d'interaction auquel chaque segment correspond a été élaborée. Une liste contenant une occurrence de chaque segment commun anonymisé a été chargée dans le modèle de langage ChatGPT[4] avec le prompt suivant, inspiré du script des fraudes amoureuses de Whitty (2015) et dont voici la version finale :
« Je vais te transmettre un fichier. Mon objectif est d'analyser et de classer chaque segment en fonction de son type de discours, en ajoutant deux colonnes : une pour la "catégorie" et une autre pour le "registre" du segment. Utilise ces catégories :
Formules de politesse : Bonjour, bonsoir, comment vas tu, bonne nuit, etc....

- *Séduction : Discours visant à flatter ou attirer l'attention de la victime, déclaration d'amour.*
- *Demande de soutien : Appels émotionnels sollicitant de l'aide, manipulation, demande financière, champ lexical de l'argent, banque, finance, donnée chiffrée.*
- *Agressivité/menace : Discours coercitif ou intimidant, discours à formulation négative.*
- *Persuasion : Tentative de convaincre la victime d'un point de vue ou d'une action spécifique, avec des arguments rationnels ou émotionnels.*
- *Complicité : questions sur le passé, les centre d'intérêt de la victime, sur la victime elle même, exemple "as tu des enfants?" "quel est ton signe astrologique?", "aimes tu la musique?".*
- *Promesse : Offres de récompenses ou d'avenir meilleur en échange de coopération.*
- *Justification : Explications pour légitimer une action ou demande douteuse.*
- *Excuses : Discours visant à s'excuser pour des erreurs.*
- *Urgence : Discours visant à faire réagir rapidement, en suscitant la panique ou l'impatience.*
- *Vie quotidienne : Questions relevant du quotidien, journée normale, travail, etc..*

Renvoie moi un fichier csv en sortie dont les colonnes sont délimitées par point-virgule».

[1] La lemmatisation consiste à réduire les mots à leur forme canonique ou de base (par exemple, transformer "marchait" en "xmarche.
[2] Une forme unique désigne une occurrence d'un mot exactement tel qu'il apparaît dans le texte, sans modification ni regroupement avec d'autres variantes (par exemple, les mots "marcher", "marchait" et "marcheront" sont considérés comme des formes uniques distinctes
[3] https://spacy.io/universe/project/textacy
[4] https://chatgpt.com, modèle GPT-4o

La définition des catégories repose sur les étapes définies par Whitty, sans les reprendre de manière directe. Les concepts ont été synthétisés et reformulés afin de les rendre opérationnels dans une démarche de classification de contenu par l'IA. Les catégories Séduction, Demande de soutien, Justification, Urgence, Persuasion, Promesse, Excuse et Complicité ont ainsi été formulées à partir du cadre fourni par Whitty (2015). Les catégories Agressivité / Menaces, Formule de politesse et Vie quotidienne ont été ajoutées par l'opérateur et correspondent à un cadre conversationnel plus courant.
Le prompt a dû être affiné à plusieurs reprises afin de limiter les erreurs de classification. Il a ensuite été demandé d'attribuer une catégorie à chaque segment afin de limiter le nombre de segments non classés.
Le modèle de langage a été exploité comme un outil pour supporter l'analyse. En effet, à la suite du traitement réalisé par le modèle, un affinage des catégories a été nécessaire. Une catégorie « Ponctuation/Emoji » est ajoutée par l'opérateur dans le prompt et chaque segment est passé en revue afin de réattribuer manuellement une catégorie plus adéquate lorsque cela est nécessaire. La réattribution manuelle par l'opérateur est effectuée en fonction des champs lexicaux présents dans chaque segment et correspondant aux descriptions initialement définies.
Finalement, un filtre a été appliqué pour ignorer les emojis, la ponctuation, les formules de politesse ainsi que les 1, 2 et 3-grammes de mots. Ce filtrage est nécessaire, car les segments de ces catégories sont issus d'interactions courantes et pas suffisamment discriminants pour établir des liens considérés comme pertinents.
Les catégories de segments sont finalement mises en perspective avec le script de la fraude amoureuse proposé par Whitty (2015) afin de déterminer quel type de segment est employé dans quelle étape du script.

Comparaison avec les relations détectées par les informations identifiantes

Cette dernière étape est effectuée afin d'observer si les liens détectés dans la première étape par les informations identifiantes sont confirmés ou non par les liens établis avec l'analyse des segments répétés.
Les liens détectés par les segments répétés sont comparés aux liens obtenus à partir des informations identifiantes extraites des textes. Chaque lien est qualifié selon s'il correspond à un lien déjà détecté ou nouvellement détecté.

- Absent : le lien est détecté par les informations identifiantes, mais n'est pas détecté par un lien linguistique.
- Confirme : le lien est détecté par les informations identifiantes et est confirmé par un ou plusieurs liens linguistiques.
- Etend : le lien n'est pas détecté à l'aide des informations identifiantes, mais est détecté par les traces langagières.

Détection des espaces de convergence en ligne

Une fois les liens entre les cas détectés par les segments répétés, la seconde question de recherche était d'exploiter ces segments communs afin de mettre en évidence des espaces de convergence sur le web, où il est inféré que les fraudeurs se sont partagé ces segments. La détection de ces espaces d'échange a pour objectif, non seulement d'établir si des auteurs de cas ont consulté des plateformes en ligne, mais également de documenter les modalités de circulation de l'information au sein de ce système de délinquance.
Dans cette partie, le terme « modèle » désigne un ensemble de phrases disponibles sur une page web accessible en source ouverte et destinées à structurer une conversation dans le cadre d'une fraude amoureuse (voir les exemples sur le Tableau 3).

Tableau 3
Exemples de modèles de phrases disponibles en ligne

Exemple de phrases formatées
« J'ai attendu trop longtemps pour avoir des liens comme les nôtres, et je ne désire rien d'autre que toi. »
« Nos âmes en fusion recouvraient nos corps de sueur et nous nous abandonnions pour une nuit à notre fureur.»
« Chaque jour, je pense à toi, chaque nuit tu es avec moi dans mes pensées et dans mon cœur. »
« Je t'aime tendrement et pour toujours quand je pense à nous deux, partageant nos espoirs et nos rêves, nos joies, nos peines, je sais que le " nous "n'aura jamais de fin. »
« Crois-moi je t'en prie, quand je te dis que je ne te quitterai jamais tu représentes trop pour moi ! »

Ces modèles, disponibles en ligne, impactent l'évaluation des relations détectées par des segments répétés. Leur présence suggère la détection de groupes de fraudeurs exploitant ces modèles plutôt que la détection d'un auteur unique. Une extraction des trois premières pages renvoyées lors d'une recherche sur le moteur de recherche Google[5] avec les mots clés « blog mougou » et contenant des modèles a été réalisée afin de pouvoir effectuer une comparaison entre les phrases types proposées sur ces sites et les segments communs détectés entre les cas d'étude lors de l'analyse correspondante.
Une nouvelle recherche sur Google a ensuite été effectuée en utilisant les segments communs tirés de ces trois pages comme mots clés (appelée itération 1 dans la suite). Cette seconde itération a pour objectif de détecter les modèles qui n'étaient pas présents dans ceux proposés au départ et d'ainsi détecter de nouveaux espaces d'échange.
Les URL détectées ont ensuite été classées en quatre catégories sur la base de termes présents dans l'adresse[6]. Le choix d'intégrer « blogspot », «*Nom Host_1*» et « eklablog » pour classer les blogs liés aux fraudeurs a été fait après avoir constaté dans les recherches manuelles qu'il s'agissait des plateformes les plus couramment utilisées pour le partage.

[5] Résultats au 24.11.2024

[6] Formule Excel pour le classement :
=SI(OU(NB.SI(C2;"*mougou*")>0;NB.SI(C2;"*Nom Host_1*")>0;NB.SI(C2;"*eklablog*")>0;NB.SI(C2;"*blogspot*")>0;NB.SI(C2;"*format*")>0;NB.SI(C2;"*formate*")>0;NB.SI(C2;"*mogo*")>0;NB.SI(C2;"*confiance*")>0);"blog mougou"; SI(OU(NB.SI(C2;"*facebook*")>0;NB.SI(C2;"*twitter*")>0;NB.SI(C2;"*tiktok*")>0;NB.SI(C2;"*reddit*")>0);"réseau social"; SI(OU(NB.SI(C2;"*leparisien*")>0;NB.SI(C2;"*lejdd*")>0;NB.SI(C2;"*leparisien*")>0;);"journal"; SI(OU(NB.SI(C2;"*tinder*")>0;NB.SI(C2;"*meetic*")>0;NB.SI(C2;"*bumble*")>0;NB.SI(C2;"*okcupid*")>0);"site de rencontre"; "autre"))))

Les URL labellisées « blog mougou » (N = 220) ont été extraites afin de récupérer un échantillon plus large de modèles de phrases. Chaque page récoltée par ce biais a été consultée manuellement, afin de filtrer celles qui ne correspondaient pas à la recherche.
À chaque page est attribué un identifiant de modèle afin de pouvoir détecter quels segments sont présents dans quels modèles. Lorsque les pages web étaient les mêmes, un même identifiant de modèle a été attribué. Afin d'assurer la complétude de l'analyse, l'opération est réitérée en utilisant les segments communs entre les cas d'étude détectés lors de la première phase de l'analyse et répondant aux critères suivants : (1) Présent dans les catégories Séduction et complicité ; (2) Contenant plus de dix mots ; (3) Ne contenant pas de phrases trop spécifiques à un cas (Dénomination, situation particulière) ; (4) Ne portant pas sur des demandes de photos. Les cas sont finalement regroupés en fonction des segments issus de modèles qu'ils contiennent, et associés aux plateformes correspondantes.

Résultats

Les résultats sont présentés selon les trois étapes de l'analyse : la détection de liens par les informations identifiantes, puis par les segments répétés (H1), et finalement la détection des espaces de convergence à partir de ces segments répétés (H2).

Détection de relations par les informations identifiantes

L'analyse des informations identifiantes a permis de mettre en évidence uniquement 18 liens entre les 180 cas de fraudes. Ceux-ci ont été catégorisés en fonction du type d'information ayant permis de faire le lien entre deux cas : information bancaire (IBAN, bénéficiaire, adresse Bitcoin), information de communication (adresse mail, numéro de téléphone), patronyme (nom et prénom) et information du profil (nom, prénom et lieu de résidence).
À noter que plusieurs liens sont observés entre de mêmes cas. Ainsi, dix relations distinctes ont été détectées entre 16 cas. Par exemple, deux cas concentrent cinq de ces dix-huit liens. Si l'on regarde plus précisément le type d'information ayant permis d'établir des relations, il est constaté qu'un seul lien est établi par l'adresse électronique fournie par le fraudeur et deux liens avec un numéro de téléphone. Les données financières, quant à elles, permettent d'établir huit liens, notamment une adresse Bitcoin commune entre deux cas. À l'exception de deux cas qui n'ont que le nom de bénéficiaire en commun, l'ensemble des relations comprennent un nom de bénéficiaire et un IBAN.
Finalement, les liens de type « PROFIL » détectés à partir des captures d'écran issues des profils des fraudeurs sur les sites de rencontre permettent de connecter trois paires de cas.
Il s'agit là des seuls liens établis à partir des informations extraites des textes lors de l'anonymisation. Moins de 10% du total des cas (16 sur 180) sont ainsi liés.

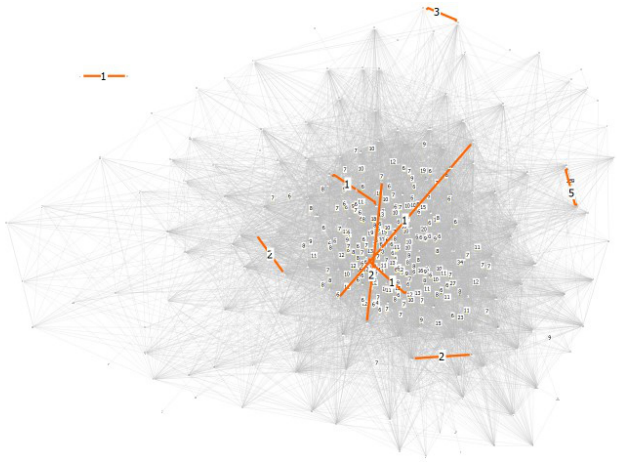
Détection de relations par les segments répétés

Avant filtrage, 10454 liens ont été détectés par 750 segments uniques. Au total 165 cas sont liés à au moins un autre cas par un segment commun (Figure 1).

Lors de l'analyse détaillée de ces segments communs, les éléments suivants ont pu être observés :

- Vingt segments sont des emojis ou de la ponctuation et établissent 4134 liens.
- 112 segments uniques (pour 2049 liens) sont des formules de politesse pour engager une conversation comme « cc ca va », « Bonjour », « comment vas-tu ? », etc. Ces segments lient 137 cas, mais leur manque de spécificité ne permet pas de les considérer comme des liens pertinents et discriminants.
- Certains segments, nettement plus longs et pouvant inclure plusieurs phrases, sont reproduits à l'identique par différents fraudeurs. Il apparaît peu plausible que de telles formules soient générées de façon spontanée et identique par des individus distincts.

Figure 1
Représentation des liens détectés à la suite de la première extraction des segments (gris, combinés aux liens établis sur les informations identifiantes (orange)



Catégorisation des segments et filtrage des relations

Sur les 1553 segments uniques pertinents catégorisés avec ChatGPT, 747 ont été reclassés lors de l'analyse manuelle (48%). Les résultats sont présentés dans le Tableau 4. Le modèle de langage a classé 63% des segments dans la catégorie « Vie Quotidienne » et n'a pas détecté d'expression d'agressivité ou de menace. Lors de la classification manuelle, les expressions destinées à mettre de la pression sur la victime, telles que « tu ne veux plus me parler ? », « Pourquoi tu me fais ça ? » ou utilisant le champ lexical de la haine (« tu me détestes ») sont classées comme « agressives ». L'IA n'a pas non plus attribué la catégorie « Persuasion ». 57% des segments « Excuse » ont néanmoins été classés correctement.

Tableau 4
Classification des segments par ChatGPT et reclassification manuelle

Catégorie	ChatGPT	Manuelle
Agressivité/Menace	0	100
Complicité	8	200
Demande de soutien	119	149
Excuses	11	19
Formule de politesse	66	130
Justification	40	14
Persuasion	0	42
Ponctuation/Emoji	0	23
Promesse	1	41
Séduction	305	244
Urgence	20	58
Vie Quotidienne	983	533

À la suite du filtrage des catégories « Formule de politesse » et « Ponctuation / Emoji », ainsi que des 1, 2 et 3 grammes, 1462 liens établis par 410 segments répétés sont conservés. Ces segments relient 98 cas distincts, soit 59% des cas détectés initialement (voir Figure 2).

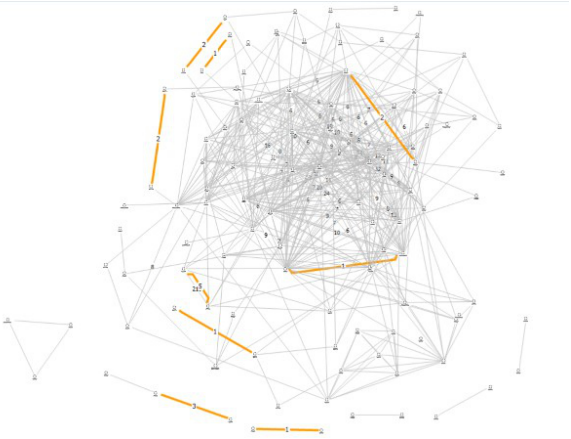
Recherche de modèles de conversation et détection des espaces en ligne de partage

Lors de la première itération de la récolte (Tableau 5), 5117 URL distinctes ont été détectées et collectées à partir de 357 segments. À noter qu'une même URL a pu être détectée à l'aide de mots clés différents (voir Tableau 8). Le nombre d'occurrences d'une même URL varie entre 1 et 184. L'URL la plus fréquemment retrouvée renvoie à un blog du domaine « blogspot.com », qui correspond au premier lien retourné lors d'une recherche Google (observation au 24.11.2024). Ce blog regroupe plusieurs dizaines de modèles de phrases.

Tableau 5
Nombre et type d'URL récoltées à l'aide des modèles de phrase (Itération 1)

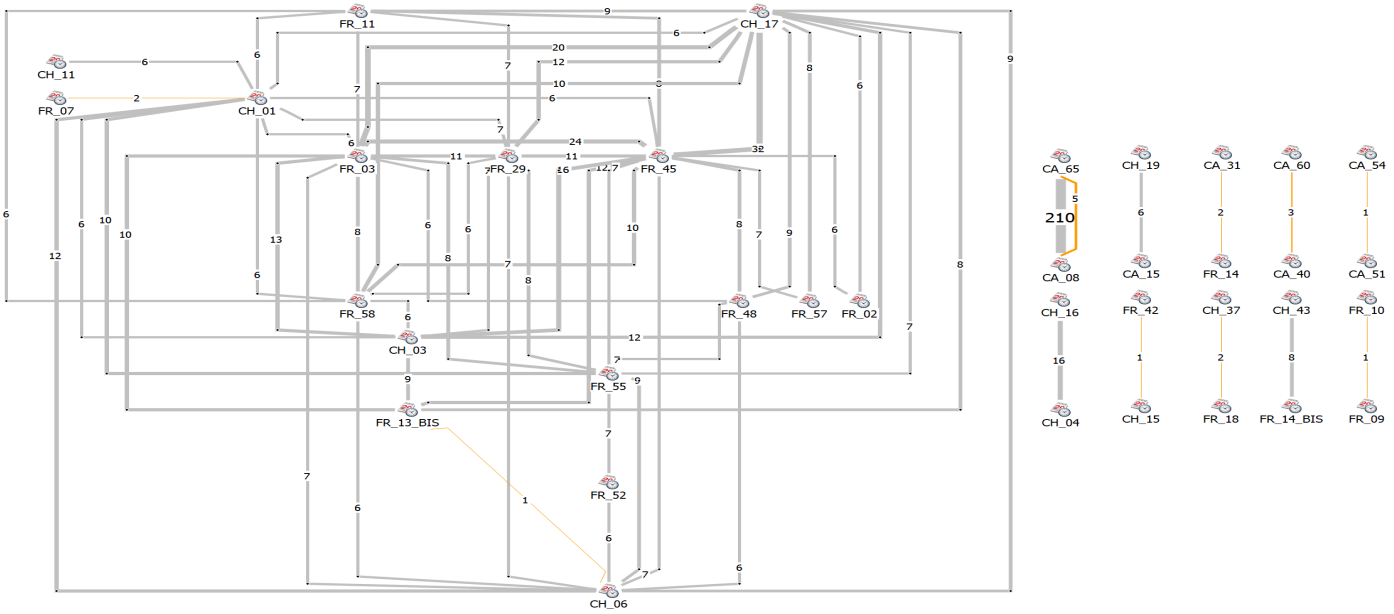
Catégorie	Nombre d'URL total	Nombre d'URL distincte
site de rencontre	11	4
journal	14	13
réseau social	915	649
blog mougou	1 380	220
autre	5 041	4208
TOTAL	7361	5094

Figure 2
Représentation des liens détectés à la suite du filtrage des segments (gris), combinés aux liens établis sur les informations identifiantes (orange)



22 cas sont liés par plus de 5 segments répétés. Les 526 liens individuels établis sont répartis entre trois catégories : « Vie quotidienne » (n=15), « Séduction » (n=20), et « Complicité » (n=491).

Figure 2.1
Ensemble des cas liés avec au moins 6 liens sur les segments communs (gris) et liens sur les infos identifiantes (orange)



La catégorie « autre » regroupe les URL n'ayant pas pu être classées. Il s'agit majoritairement de sites dédiés à des conseils sur les rencontres amoureuses ou à des magazines. Un autre blog ressort particulièrement avec les mots clés « format » et « mougou ».
Les réseaux sociaux incluent Facebook, TikTok, Twitter et Reddit. Un profil Facebook particulier a été détecté 124 fois, soit par 124 combinaisons de mots clés. Après consultation, il est constaté que les publications de cette page se composent uniquement de modèles de phrases destinées aux fraudes amoureuses. Au 04.04.2025, la page existait toujours, mais toutes les publications ont été supprimées.
Sur les 220 pages dédiées au « mougou », 39 ont été conservées pour leur pertinence et ont pu être extraites automatiquement. 514 segments uniques ont été détectés sur ces pages, dont 64 établissent des liens entre les cas.
Lors de la seconde récolte, 8159 URL ont été détectées à partir de 188 segments utilisés comme mots clés. Après classification, l'échantillon est réduit à 250 URL. 109 pages de blogs « mougou » distinctes ont été détectées, dont 24 sont considérées comme pertinentes.
La combinaison des deux itérations a permis de mettre en évidence 301 plateformes dont ont été extraits 2483 segments, dont 705 segments uniques. Onze noms d'hôtes et trois noms de domaines sont partagés par 39 cas. Il n'est néanmoins pas possible d'inférer quel nom d'hôte exact pourrait avoir été utilisé dans un cas précis lorsque les modèles de texte se trouvent sur plusieurs plateformes différentes.

Cluster 1 : Domaine « blogspot.com »

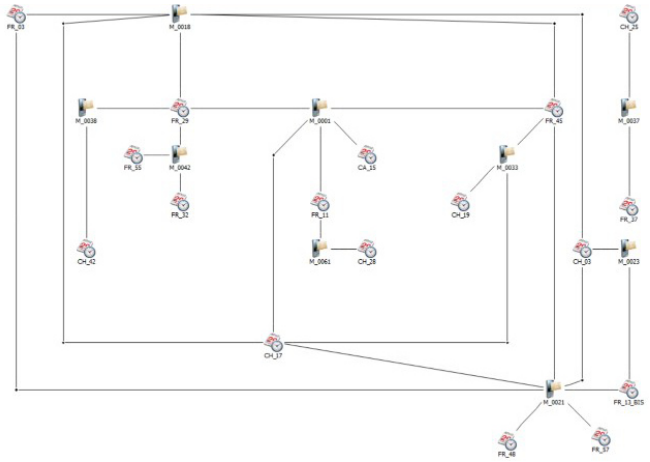
Le domaine « blogspot.com » regroupe 9 noms d'hôtes sur lesquels 68 segments répétés liés à 39 cas ont été détectés. Ces segments se retrouvent presque exclusivement sur des pages du domaine blogspot.com (n=31). Deux pages du domaine « eklablog.com » sont également liées. Sur ces 68 segments, 40 sont des phrases déclaratives et 28 sont des questions.
Neuf segments se retrouvent uniquement sur trois pages distinctes du site correspondant au « Host_1 » du domaine « blogspot.com », soit trois modèles (voir Tableau 6). Il s'agit de trois onglets distincts d'un même blog proposant chacun des modèles de conversation. Parmi les neuf segments, trois sont des questions, les six autres sont des phrases déclaratives, dont cinq commencent par la formulation « Moi je » ou « Moi j' ».

Tableau 6
Segments issus de modèles du nom d'hôte « Host_1 »

ID line	Line
M_0001	Moi j'aime la cuisine, musique, danse, resto, lecture, natation et aussi j'adore les ballades familiales
M_0018	Moi j'adore les ballades en bordure de l'eau et aussi le cinéma, théâtre, plage, découvertes des nouveaux horizons et pays
M_0021	Moi je n'ai aucune idée pour l'instant, mais c'est de rendre la femme heureuse et aussi d'être heureux pour l'instant ce sont mes projets on verras pour le reste dans l'avenir
M_0023	Comme au cœur de nos rapports Il existe entre nous un respect et une appréciation profonde Notre relation ne peut que s'affermir et embellir chaque jour si la volonté d'aimer et la confiance y existe entre nous.
M_0033	Moi je les aimes bien, mais j'en ai pas
M_0037	Tu as déjà eu des relations vague ou quelconque après ta séparation?
M_0038	Moi je suis nouveau et tu es la première et j'espère bien la dernière
M_0042	Tu crois trouver ce grand amour (l'homme parfait) ici sur le réseau social ?
M_0061	Et ben dis moi tu as des principes a respecté ou des devises pour t'encourager dans tout ce que tu entreprend ?

Ces segments sont partagés par 17 cas (Figure 3). Les cas les plus connectés sont FR_45 et CH_17 avec quatre segments communs.

Figure 3
Représentation des cas et des segments issus de modèles du nom d'hôte « Host_1 »



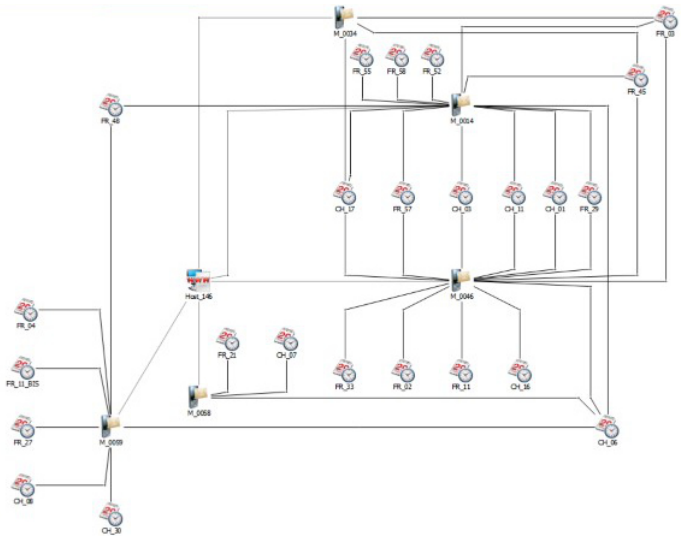
Cluster 2 : Domaine « eklablog.com »

Pour ce domaine, un seul blog a été retrouvé, nommé ici « Host_16 ». Cinq cas et deux segments sont liés au modèle : « Dis moi pour toi le mariage est très important ? » et « Je crois que non, mais le plus important pour moi c'est de trouver l'âme sœur ». L'un des segments est une question, l'autre une phrase déclarative. A noter que ces segments sont également en ligne sur cinq autres noms de domaine.

Cluster 3 : Domaine « gov.gh »

Pour ce nom d'hôte, une seule URL a été détectée. Ce domaine peut surprendre, dans la mesure où le domaine de second niveau « .gov » renvoie généralement à un site gouvernemental du pays concerné, en l'occurrence le Ghana. En l'espèce, il semble toutefois s'agir d'un site communautaire proposant un espace de forum, dont est issu le modèle détecté comprenant cinq segments liés à 24 cas. Le cluster contient uniquement des questions : « Tu fume ? Tu bois ? », « Quels types de sports et loisirs pratiques-tu ? », « Quels sont tes loisirs ? », « Tu vie seul ? Ou en famille ? », « Tu as des enfants ? ». A noter que les segments qui le composent apparaissent également sous neuf autres noms d'hôtes.

Figure 4
Représentation des cas et des segments issus de modèles du cluster 3



Analyse des segments répétés non liés à des modèles présents en ligne

Les segments communs non liés à des modèles présents en ligne ont été analysés plus en détail afin d'affiner la caractérisation des critères de pertinence des liens établis. La majorité des segments communs non attribués à des modèles sont classifiés en « Vie Quotidienne » et regroupent tout ce qui est lié aux références au travail, à la vie en général (Tableau 7). Certaines phrases sont dérivées des questions proposées dans les modèles. Les 85 segments « Complicité » peuvent être subdivisés en trois catégories. 41 segments classés en « Complicité » sont des questions, ce qui représente presque la moitié d'entre eux (49%). 20 segments commencent par une formulation du type « moi je », « moi j' » ou « moi », ou la contiennent. Ces segments sont structurés comme une réponse à une question posée précédemment, et comportent une dimension descriptive de la personnalité fictive du fraudeur.

Parmi les autres segments déclaratifs, sept ont moins de 12 mots, les autres sont de longues tirades mettant en avant ce que recherche le personnage joué par le fraudeur dans une relation, ou exprimant ses craintes.

La différence majeure entre les segments de type « Complicité » et « Vie Quotidienne », est que les premiers portent sur les préférences de la victime, sa vision du couple, ce qu'elle recherche. Les questions sont orientées sur le développement de la relation et ont pour objectif d'apprendre à connaître la victime / le personnage fictif du fraudeur. Les questions dans les segments « Vie Quotidienne » sont de type « Tu fais quoi ? » « Tu es au travail ? », et portent sur l'instant présent, la connexion entre les deux membres de la relation à l'instant où la conversation a lieu. Les sujets principaux sont le travail, le déroulement de la journée, l'activité de la personne au moment de la conversation.

Tableau 7
Catégorisation des segments communs

Catégorie	Total distinct de segments
Vie quotidienne	134
Complicité	85
Demande de soutien	36
Séduction	31
Persuasion	7
Urgence	6
Promesse	3
Excuses	3
Agressivité/Menace	3

Discussion

La première hypothèse proposait de tester si l'analyse des segments répétés entre des cas de fraude amoureuse permettait d'établir des relations entre ces cas. Pour cela, deux étapes ont été effectuées. Tout d'abord, afin d'avoir un premier aperçu des liens existant, les informations identifiantes présentes dans les textes ont été comparées. Ces textes ont ensuite été analysés afin de mettre en évidence les segments communs à deux ou plusieurs cas. Les deux types de liens ont finalement été mis en commun afin de vérifier si les liens établis par les informations identifiantes étaient confirmés ou non par les segments répétés, et si ces segments permettaient de mettre en évidence de nouveaux liens non établis au préalable par les informations identifiantes.

Liens sur les informations identifiantes

Sur 180 cas présents dans le corpus de départ, seuls 18 ont présenté des liens basés sur les informations identifiantes disponibles avec d'autres cas, soit 10%. Cela peut sembler faible, compte tenu du nombre d'informations à disposition. Cependant, ce type de donnée peut facilement être modifié d'un cas à l'autre. Il est en effet très simple de créer un patronyme ou une adresse électronique différente, par exemple. Les variations dans les numéros de téléphone peuvent également s'expliquer par la facilité d'accès à des cartes SIM prépayées selon la localisation ou l'achat de numéros de téléphone temporaires sur internet. Une grande diversité de banques et de numéros de comptes bancaires utilisés a également été constatée, ainsi que de pays d'origine des banques concernées. Selon les entretiens menés par Cretu-Adatte, Zbinden, *et al.* (2024), de nombreux intermédiaires entrent en jeu entre le moment où l'argent est versé par la victime et le moment où l'auteur récupère les gains de la fraude. Il peut s'agir de personnes issues de l'entourage du fraudeur, comme des voisins ou des amis acceptant de prêter leur compte, ou des membres du réseau ayant pour rôle d'effectuer les transferts intermédiaires d'argent, moyennant un pourcentage sur les gains. Des mules peuvent également être recrutées pour assurer les transferts. Ceci peut expliquer la diversité de comptes observée.

Liens sur les segments répétés

La proportion des liens établis par les informations identifiantes est très faible par rapport aux liens détectés sur les segments répétés, même après filtrage (10454 segments avant filtrage contre 1462 après filtrage). Cependant, la majorité de ces segments traitent de sujets liés à la vie quotidienne ou à des tentatives d'établir une complicité avec la victime. Les liens les plus intéressants peuvent être ceux de la catégorie « Demande de soutien », mais 35 des 36 segments communs détectés sont issus des deux cas, dont le lien a été détecté par les informations identifiantes. Ce lien peut être considéré comme confirmé.

Sur l'ensemble des liens, 216 sont détectés par les informations identifiantes et des segments répétés. 210 lient deux cas. Cinq segments en relient deux autres et quatre relient deux cas également. Tous les autres liens contribuent à étendre les séries, ce qui signifie que seuls les segments répétés permettent de relier les cas entre eux. 98 cas sur les 180 sont concernés. La majorité des liens sont constitués de phrases de la vie quotidienne ou de segments issus de modèles. Ainsi, ces phrases sont pour la plupart très génériques et pas suffisamment spécifiques pour être considérées comme des liens forts. Cela montre une uniformité globale dans l'ensemble du corpus et dans les échanges utilisés par les fraudeurs, ce qui rend difficile de détecter les caractéristiques distinctives permettant d'établir des liens suffisamment individualisés entre les différents cas.

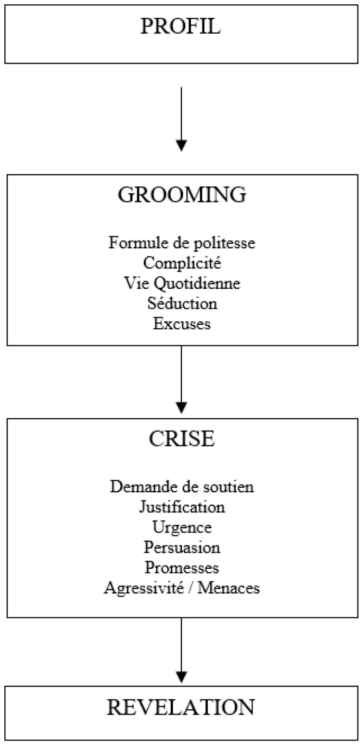
Classification des segments

La classification des segments répétés avec ChatGPT a montré une efficacité moyenne et a nécessité une reclassification de 52% des segments. Le modèle de langage n'a notamment pas été capable de reconnaître de segment associé à la catégorie « Agressivité et menaces ». Cela peut être dû au fait que les segments sont envoyés à ChatGPT sans aucun élément de contexte, ce qui peut complexifier l'interprétation. La surreprésentation de la catégorie « Vie Quotidienne » (983 contre 533 par classification manuelle) pourrait être expliquée par le fait que, dans le prompt, il était demandé d'effectuer une classification systématique de chaque segment afin d'éviter une catégorie « Inconnu ». On peut alors émettre l'hypothèse que la catégorie « Vie quotidienne » manquait de précision dans sa définition lors de l'envoi du prompt, et a pu être associée aux segments que ChatGPT n'a pas su classer dans une autre catégorie.
À noter que la classification comporte malgré tout une part de subjectivité. En effet, la perception de la menace ou de l'agressivité d'un message, ou la nuance entre « Complicité » et « Vie Quotidienne » peuvent être difficiles à saisir pour un modèle de langage, mais risque également de varier selon l'opérateur lors d'une classification manuelle. Il est ainsi important d'adopter une systématique de classification claire et précise afin de mitiger les risques d'intervariabilité et d'intravariabilité. La conclusion de cette expérience est que la classification de segments de textes isolés de leur contexte avec ChatGPT comme modèle de langage peut être exploitée en tant qu'outil de triage, mais pas en tant que méthode de classification complètement valide et reproductible. Dans ce contexte, un retour au texte par un opérateur humain est nécessaire afin de réattribuer les segments à la bonne catégorie lorsqu'ils ont été mal classés. Le gain de temps a cependant été significatif dans le cadre de cette étude par un regroupement initial facilitant les lectures et les comparaisons. Dans de futurs travaux, il serait utile de réévaluer la classification par un modèle de langage en intégrant les messages précédents et suivant le message à classer, afin d'étudier l'influence du contexte sur l'interprétation du texte. Il serait également pertinent de faire effectuer la vérification à plusieurs opérateurs humains indépendants. La démarche n'était pas possible dans le cadre de cette étude.

Cadre discursif

Le script décrit dans la revue de la littérature ainsi que les modèles de conversation relevés par Anesa (2020) et Faber (2024) permettent d'émettre l'hypothèse d'une macro et d'une micro structure (Gledhill *et al.*, 2017) spécifiques au cas des fraudes amoureuses. La macrostructure correspond aux grandes étapes du script. Pour chaque étape, les modèles de conversation forment des microstructures de dialogues qui varient suivant les sources choisies (Figure 5).
En d'autres termes, les conversations évoluent dans un cadre déterminé, pouvant être associé à la notion de « moule discursif » (Gautier, 2018). Pour Bach *et al.* (2022), un texte doit être considéré comme une unité de communication qui ne peut être compris indépendamment du contexte dans lequel il est produit : « que ce soit au niveau communicationnel, pragmatique, sémantique ou encore psychologique, cognitif et social, le texte est bien une structure stabilisée dans le temps et l'espace, reconnaissable, unique sans être tout à fait autonome, et ayant une fonction identifiée et reconnue par les individus dans la communication humaine » (p.13).

Figure 5
Répartition des catégories de segments répétés en fonction des étapes de la fraude amoureuse décrite par Whitty (2015)



Ainsi, les segments communs détectés se répartissent entre deux étapes du script : le Grooming et la Crise. Cela s'explique, car la création du profil n'implique pas de conversation, et que les conversations n'allaient jamais au-delà des demandes d'argent. Ainsi, la nature des données d'analyse induit un biais inhérent à la pratique du leurrage d'escrocs. Il est important de relever également que la démarche initiale de collecte (Zbinden *et al*., 2023) visait un objectif de récupération d'informations utiles à la définition de stratégie de perturbation, les messages orientés vers cet objectif, pouvaient ainsi conditionner les réponses des fraudeurs et ne pas toujours correspondre aux comportements de personnes en recherche d'une relation amoureuse.

Si les conversations textuelles issues des fraudes amoureuses suivent un cadre qui leur est propre, des modèles ou « moules » reconnaissables issus des processus d'échanges d'expérience entre les fraudeurs s'y intègrent. Dans cette optique, les modèles extraits des plateformes d'échange en source ouverte contribuent au niveau micro à structurer le cadre textuel (moule discursif) au niveau macro. Ainsi, le fait que ces fraudeurs partagent ces modèles de conversation, un acte en soi non répréhensible, dans le but d'aider leurs confrères à hameçonner leurs victimes, constitue une forme observable (Tremblay, 2010) de ce système de délinquance lié aux fraudes amoureuses. Comme l'exprime Cusson (2005), « la fonction première d'un système de délinquance est d'assurer l'impunité de ceux qui le maîtrisent : ceux-ci ayant trouvé le moyen de tromper les victimes, de neutraliser les dispositifs de protection ou de déjouer la police. » (p. 94).

A noter que seuls trois types de modèles de phrases sont retrouvées en ligne : Vie quotidienne, complicité et séduction). Ces trois catégories peuvent être associées dans la structure globale de Whitty (2015) à l'étape de « Grooming », c'est-à-dire au développement de la fausse relation amoureuse. Aucun modèle n'a été classé dans une catégorie pouvant correspondre aux étapes ultérieures, notamment les demandes d'argent ou de service. Cette observation permet d'émettre l'hypothèse que les échanges d'information entre fraudeurs se font majoritairement pour favoriser l'étape de séduction, mais que la structuration des demandes d'argent est plus spécifique à la situation de la victime et au scénario mis en œuvre. Les résultats confirment les observations de Whitty et Buchanan (2012) qui mettent en avant les variations temporelles entre les cas, indiquant des périodes de « Grooming » plus ou moins longues selon la capacité du fraudeur à hameçonner sa victime. Les demandes d'argent interviennent alors dans des contextes et des temporalités différentes selon les cas, poussant le fraudeur à une certaine adaptabilité dans son cadre discursif. Néanmoins, il convient de relever que les conversations n'allaient jamais au-delà des demandes d'argent.

Ainsi l'inférence d'une relation entre des cas fondées sur la reconnaissance de segments répétés extraits de leur contexte d'énonciation semble devoir être considérée avec prudence. La trace langagière repose sur l'idée que la trace textuelle porte des indices sur son auteur et sur le cadre d'activité dans lequel elle est produite. Toutefois, lorsque ces segments sont traités comme des unités indépendantes, leur interprétation peu perdre une partie de sa robustesse. La présence d'un segment répété n'est pas nécessairement synonyme de relation directe entre deux cas, mais peut refléter la standardisation des schémas interactionnels propres à ce type de fraude. Une perspective plus contextualisée, intégrant la dynamique conversationnelle, les séquences adjacentes ou les actes de discours associés, semble nécessaire pour renforcer la validité inférentielle des rapprochements linguistiques.

Liens fondés sur des modèles présents en ligne

Dans le cadre de cette étude, 357 segments répétés détectés lors de l'analyse ont conduit à l'identification de 301 plateformes, à partir desquelles un ensemble élargi de 2 483 segments a été extrait, dont 705 segments uniques sont retrouvés dans les cas. Ces résultats viennent confirmer l'hypothèse 2, selon laquelle les segments répétés partagés entre plusieurs cas peuvent servir à détecter des espaces de convergence, entendus comme des espaces d'échange entre fraudeurs. La démarche permet ainsi de reconnaître les segments de texte issus de moules discursifs partagés, autrement dit des modèles potentiellement repris par plusieurs fraudeurs différents.

Globalement, cinquante-trois pour cent des liens sur les segments répétés ont pu être associés à des modèles de textes accessibles en ligne (784 sur 1462 liens). Trois clusters fondés sur des segments issus de noms de domaine identiques ont pu être mis en évidence. Cette observation suggère un usage partagé des plateformes rattachées à ces domaines. L'interprétation doit toutefois rester prudente, en particulier lorsqu'un même modèle est présent sur plusieurs plateformes distinctes, puisqu'il devient alors impossible de conclure à l'usage commun d'une plateforme précise.

L'analyse des cas mobilisant ces modèles a néanmoins permis de mettre en évidence trois URL relevant d'un nom d'hôte unique, au sein duquel plusieurs segments n'ont été retrouvés que dans cet espace. Neuf modèles, liés à 17 cas, y sont présents. Bien que l'existence d'autres espaces de partage en ligne ne puisse être exclue, la pertinence du lien pourrait paraître plus forte. Néanmoins, le site comporte dix onglets (dont trois correspondent aux URL détectées) qui contiennent 4848 propositions de phrases. En comparaison avec les autres sites détectés, la communauté de l'espace semble très active, ce qui peut impacter sa visibilité. En effet, il s'agit du premier résultat proposé par Google lors des recherches. Ainsi, si l'usage d'un modèle accessible en ligne renseigne sur une pratique de partage ou de consultation d'un même espace numérique, il ne saurait être interprété formellement comme un signe d'appartenance à une même communauté ou à un même système de délinquance.

Variations diachroniques et obsolescence des scripts à l'ère de l'IA

Un enjeu interprétatif tient au caractère évolutif des scripts employés par les fraudeurs. Les pratiques discursives observées à un moment donné peuvent devenir obsolètes sous l'effet de multiples facteurs : circulation de nouveaux modèles, adoption de techniques issues d'autres régions ou d'autres formes de fraudes, ajustements consécutifs à des actions de perturbation ou à la visibilité médiatique de certains procédés. Les plateformes de partage identifiées dans cette étude (blogs, forums, réseaux sociaux) fonctionnent comme des espaces d'innovation discursive, où des reformulations et adaptations sont partagées. Ces variations soulèvent un défi méthodologique : un corpus collecté en 2021 reflète un état particulier du système de délinquance, qui peut diverger sensiblement des pratiques actuelles. Une mise à jour régulière des bases de segments, voire la constitution d'une chronologie des modèles circulants, pourrait permettre de suivre l'évolution des scripts et des « moules discursifs ».

A cet égard, il apparaît essentiel d'intégrer l'impact croissant de l'IA générative dans l'analyse des fraudes (Asyalı *et al*., 2026; Leong *et al*., 2024). Les effets de l'usage de l'IA générative par les fraudeurs sur la pertinence des relations inférées par la similitude de discours pourraient impliquer de repenser les méthodes. En ce sens, le passage progressif d'un système fondé sur des scripts statiques échangés à un système fondé sur des modèles génératifs pourrait marquer des transformations des systèmes de délinquance, impliquant la nécessité de nouvelles approches : détection stylistique d'IA, analyse de patterns conversationnels non-humains, ou identification de signatures computationnelles propres à certains modèles utilisés par les fraudeurs.

Conclusion

Les communications issues des fraudes amoureuses permettent de détecter des informations identifiantes, comme des pseudonymes, des numéros de téléphone ou des identifiants de compte bancaire utiles à la détection de relations entre les cas. En effet, afin de pouvoir communiquer, les fraudeurs doivent transmettre des identités, des moyens de contact et des moyens de paiement lorsqu'une transaction est demandée au cours des échanges avec la victime. Néanmoins, s'il peut sembler aisé d'établir des liens entre les différents cas analysés à l'aide de ces informations, cette étude a mis en évidence une importante diversité d'identifiants entre les cas. Cette variabilité semble pouvoir s'expliquer par une certaine facilité de changer les identités exploitées tant pour la construction des profils que pour les paiements reposant sur l'usage d'intermédiaires financiers. Ainsi, sur les 180 cas de fraudes étudiés, seuls 18 liens, correspondant à dix relations distinctes entre 16 cas, ont été détectés (soit moins de 10% des cas).

L'analyse linguistique des textes écrits par les fraudeurs semble ainsi constituer une voie de recherche prometteuse pour détecter des similitudes utiles à la reconstruction des séries et la compréhension des systèmes de délinquance. Ainsi, cette étude a démontré comment une approche fondée sur l'analyse de segments de texte répétés permet de détecter des similitudes entre les conversations. Un nombre très important de similitudes qui semblent peu discriminantes sont observées. En effet, des phrases courtes, des emojis, de formules de politesse ou de texte ayant trait à la vie quotidienne sont très souvent commun entre les textes. Néanmoins, l'analyse des segments répétés a permis de mettre en évidence des phrases plus longues et complexes communes à certains cas. De telles similitudes peuvent constituer l'indice de l'activité d'un même fraudeur ou d'un groupe de fraudeurs.

Afin d'évaluer la pertinence des relations, un travail de catégorisation, d'abord assisté par ChatGPT puis repris manuellement, a mis en évidence les limites du modèle de langage pour l'interprétation fine de certains segments, en particulier ceux relevant de la persuasion, de l'agressivité ou de la complicité. Après filtrage des catégories peu discriminantes ainsi que des segments trop courts, 1 462 liens fondés sur 410 segments répétés ont été conservés, reliant 98 cas distincts. Parmi ceux-ci, 22 cas présentent plus de cinq segments répétés. 216 liens fondés sur des segments répétés ont pu être considérés comme confirmés, c'est-à-dire également liés par des informations identifiantes. À noter que 210 d'entre eux sont observés entre deux cas. Les autres liens détectés par les segments étendent les séries.

L'analyse révèle en outre l'existence de pratiques de partage en ligne de textes « modèles », utilisés par les fraudeurs pour structurer et faciliter leurs interactions avec les victimes. Par conséquent, les similitudes fondées sur des phrases retrouvées sur des pages accessibles en ligne, désignées comme des « segments issus de modèles », qui représentent 53 % des relations détectées, appellent une interprétation prudente.

Il n'en demeure pas moins que les traces langagières issues des échanges ouvrent une autre perspective d'analyse du phénomène, en permettant d'appréhender l'écosystème de partage en ligne par l'identification d'espaces de convergence entre auteurs, tels que des blogs ou des forums. En effet, les segments caractéristiques de la fraude ont été exploités lors de recherches en ligne pour détecter ses espaces et reconstruire des sous-groupes de cas les exploitant. La détection de 301 plateformes et de plusieurs clusters confirme l'intérêt de cette approche pour documenter l'écosystème numérique dans lequel s'inscrivent ces pratiques. Cette interprétation doit néanmoins rester prudente, dans la mesure où la diffusion d'un même modèle sur plusieurs sites et la forte visibilité de certains espaces limitent la portée inférentielle des rapprochements observés.

L'analyse des segments répétés non liés à des modèles accessibles en ligne apporte un éclairage complémentaire. Ces segments relèvent majoritairement des catégories « vie quotidienne » et « complicité ». Les premiers portent surtout sur le travail et le déroulement de la journée, tandis que les seconds concernent davantage la construction de la relation, les préférences de la victime et la mise en scène du personnage du fraudeur. Ainsi, au-delà de la détection d'espaces de partage, l'analyse linguistique permet d'identifier des moules discursifs et des schémas interactionnels récurrents, offrant un accès plus fin aux pratiques de communication mobilisées lors de la fraude amoureuse.

Dans l'ensemble, ces résultats montrent que les informations identifiantes semblent permettre d'établir un nombre restreint, mais potentiellement robuste de relations entre les cas de fraudes, tandis que les traces langagières offrent un potentiel heuristique bien plus large pour détecter des rapprochements entre cas, à condition de faire l'objet d'un filtrage rigoureux. En effet, si la majorité des segments répétés sont considérés comme de faible pertinence, la détection de modèles de phrases issus de plateformes sur le web permet d'émettre l'hypothèse d'un partage global de phrases « modèles ». Les segments non rattachés à de tels modèles, relevant principalement de la vie quotidienne et de la complicité, mettent en évidence des moules discursifs et des schémas interactionnels récurrents dans la construction du lien frauduleux. Les traces langagières permettent ainsi la détection des plateformes d'échange, et l'affinage de leur analyse pourrait permettre de renforcer les liens potentiels entre de nouveaux cas de fraude amoureuse.

Références

Abubakari, Y. (2024). Modelling the modus operandi of online romance fraud : Perspectives of online romance fraudsters. *Journal of Economic Criminology*, *6*, 100112. https://doi.org/10.1016/j.jeconc.2024.100112

Addawood, A., Badawy, A., Lerman, K., & Ferrara, E. (2019). Linguistic Cues to Deception : Identifying Political Trolls on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 15-25. https://doi.org/10.1609/icwsm.v13i01.3205

Adou, E. F.-S. (2022). Les brouteurs d'Abidjan : Étude socio-anthropologique d'une sous-culture juvénile déviante. *RESET*, *11*. https://doi.org/10.4000/reset.4038

Akermann, A. (2025). *Les briseurs de cœurs du numérique : Comment les « arnacœurs » s'y prennent | VZ VermögensZentrum*. Vermoegens Zentrum. https://www.vermoegenszentrum.ch/fr/competences/les-briseurs-de-coeurs-du-numerique-comment-les-arnacoeurs-sy-prennent

Anesa, P. (2020). Lovextortion : Persuasion strategies in romance cybercrime. *Discourse, Context & Media*, *35*, 100398. https://doi.org/10.1016/j.dcm.2020.100398

Asyalı, A. N., Frank, M.-L., & Hölzmer, P. (2026). Fake it till you make it : The psychological and communication tactics behind "Pig Butchering" scams. *Journal of Cybersecurity*, *12*(1), tyag003. https://doi.org/10.1093/cybsec/tyag003

Atta-Asamoah, A. (2009). Understanding the West African cyber crime process. *African Security Review*, *18*(4), 105-114. https://doi.org/10.1080/10246029.2009.9627562

Bach, M., Maazaoui, H., & Gautier, L. (2022). Construction approach as a method of textual analysis. Propositions based on a corpus of texts concerning the economic situation. *Écho des études romanes*, *18*(1), 11-27. https://doi.org/10.32725/eer.2022.002

Barnor, J. N. B., Boateng, R., Kolog, E. A., & Afful-Dadzie, A. (2020). Rationalizing Online Romance Fraud : In the Eyes of the Offender. *Association for Information Systems*. https://aisel.aisnet.org/amcis2020/info_security_privacy/info_security_privacy/21

Bilz, A., Shepherd, L. A., & Johnson, G. I. (2023). Tainted Love : A Systematic Literature Review of Online Romance Scam Research. *Interacting with Computers*, *35*(6), 773-788. https://doi.org/10.1093/iwc/iwad048

Bollé, T. (2025). *Utilisation des liens non évidents pour l'analyse de fraude en ligne* [Thèse de doctorat]. Université de Lausanne.

Bonnafous, S. (1988). *André Salem, Pratique des segments répétés. Essai de statistique textuelle—Compte Rendu*. 243-245.

Buchanan, T., & Whitty, M. T. (2014). The online dating romance scam : Causes and consequences of victimhood. *Psychology, Crime & Law*, *20*(3), 261-283. https://doi.org/10.1080/1068316X.2013.772180

Buil-Gil, D., & Zeng, Y. (2022). Meeting you was a fake : Investigating the increase in romance fraud during COVID-19. *Journal of Financial Crime*, *29*(2), 460-475. https://doi.org/10.1108/JFC-02-2021-0042

Carpentier, J., Viau-Quesnel, C., Potz, A., Lavertu, S., Marchand, C., Campbell, F., & Thibodeau, M. (2024). *Fraudes amoureuses en ligne : Comprendre les personnes victimes et leurs besoins* [Rapport scientifique final déposé au Ministère de la Justice Québec. Programme de subvention pour favoriser la recherche, l'information, la sensibilisation et la formation en matière d'aide aux victimes d'actes criminels.].

Carter, E. (2024). *The Language of Romance Crimes : Interactions of Love, Money, and Threat* (1re éd.). Cambridge University Press. https://doi.org/10.1017/9781009273008

Cretu-Adatte, C., Azi, J. W., Beaudet-Labrecque, O., Bunning, H., Brunoni, L., & Zbinden, R. (2024). Unravelling the organisation of ivorian cyberfraudsters : Criminal networks or organised crime? *Journal of Economic Criminology*, *3*, 100056. https://doi.org/10.1016/j.jeconc.2024.100056

Cretu-Adatte, C., Zbinden, R., Brunoni, L., Bunning, H., Azi, J. W., & Beaudet-Labrecque, O. (2024). How do Ivorian Cyberfraudsters Manage Their Criminal Proceeds? *European Journal on Criminal Policy and Research*, *30*(3), 359-378. https://doi.org/10.1007/s10610-024-09597-7

Cusson, M. (2005). *La délinquance, une vie choisie. Entre crime et plaisir*. *Travaux en criminologie*.

De Jong, K. (2019). *Detecting the online romance scam : Recognising images used in fraudulent dating profiles* [Department of EEMCS, University of Twente]. https://essay.utwente.nl/80084/1/Jong_de_MA_EEMCS.pdf

Degeneve, C., Longhi, J., & Rossy, Q. (2022). Analysing the digital transformation of the market for fake documents using a computational linguistic approach. *Forensic Science International: Synergy*, *5*, 100287. https://doi.org/10.1016/j.fsisyn.2022.100287

Degeneve, C., Longhi, J., & Rossy, Q. (2024). Distinguishing Sellers Reported as Scammers on Online Illicit Markets Using Their Language Traces. *Languages*, *9*(7), 235. https://doi.org/10.3390/languages9070235

Faber, P. (2024). The frames of romance scamming. *Research in Laguage*, *22*(1), 1-23. https://doi.org/10.18778/1731-7533.22.1.01

Gautier, L. (2018). *Approcher les discours spécialisés par la méta-catégorie du figement*. https://cel.hal.science/cel-01742139v1

Gledhill, C., Patin, S., & Zimina, M. (2017). Lexico-grammaire et textométrie : Identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français. *Corpus*, *17*. https://doi.org/10.4000/corpus.2868

Halliday, M. A. K. (2014). *An introduction to functional grammar*. Edward Arnold.

Huhn, C. K. (2023). *Laundering love: A multi-case analysis of the evolution of romance scam victims into co-offending money mules* [Thèse de doctorat, Naval Postgraduate School]. https://www.tandfonline.com/doi/epdf/10.1080/15564886.2021.2018080?needAccess=true

Kopp, C., Layton, R., Sillitoe, J., & Gondal, I. (2016). *The Role Of Love Stories In Romance Scams : A Qualitative Analysis Of Fraudulent Profiles*. https://doi.org/10.5281/ZENODO.56227

Kydd, M., Shepperd, L., Johnson, G. I., & Szymkowiak, A. (2024). *Love bytes : Improving romance fraud prevention*. CREST SECURITY REVIEW.

Lam, T., Demange, J., & Longhi, J. (2021, janvier). Attribution d'auteur par utilisation des méthodes d'apprentissage profond. *EGC 2021 Atelier « DL for NLP : Deep Learning pour le traitement automatique des langues »*. https://hal.archives-ouvertes.fr/hal-03121305

Lazarus, S., Whittaker, J. M., McGuire, M. R., & Platt, L. (2023). What do we know about online romance fraud studies? A systematic review of the empirical literature (2000 to 2021). *Journal of Economic Criminology*, *2*, 100013. https://doi.org/10.1016/j.jeconc.2023.100013

Lee, K.-F., Chan, M. Y., & Mohamad Ali, A. (2023). Self and desired partner descriptions in the online romance scam : A linguistic analysis of scammer and general user profiles on online dating portals. *Crime Prevention and Community Safety*, *25*(1), 20–46. https://doi.org/10.1057/s41300-022-00169-7

Leong, W. Y., Leong, Y. Z., & San Leong, W. (2024). The Intersection of scammers and artificial intelligence. *2024 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, 539–540.

Pollitt, M., Casey, E., Jaquet-Chiffelle, D.-O., & Gladyshev, P. (2018). *A Framework for Harmonizing Forensic Science Practices and Digital/Multimedia Evidence*. Organization of Scientific Area Committees for Forensic Science. https://doi.org/10.29325/OSAC.TS.0002

Reep-van Den Bergh, C. M. M., & Junger, M. (2018). Victims of cybercrime in Europe : A review of victim surveys. *Crime Science*, *7*(1), 5. https://doi.org/10.1186/s40163-018-0079-3

Renaut, L., Ascone, L., & Longhi, J. (2017). De la trace langagière à l'indice linguistique : Enjeux et précautions d'une linguistique forensique. *Ela. Études de linguistique appliquée*, (188), 423–442.

Rossy, Q., & Ribaux, O. (2020). Orienting the Development of Crime Analysis Processes in Police Organisations Covering the Digital Transformations of Fraud Mechanisms. *Eur J Crim Policy Res*. https://doi.org/10.1007/s10610-020-09438-3

Rubin, V. L. (2016). Deception Detection and Rumor Debunking for Social Media. In L. Sloan & A. Quan-Haase, *The SAGE Handbook of Social Media Research Methods* (p. 342–363). SAGE Publications Ltd. https://doi.org/10.4135/9781473983847.n21

Salem, A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, *1*(2), 5–28. https://doi.org/10.3406/hism.1986.1518

Schokkenbroek, J. M., & Snaphaan, T. (2025). Love as Bait : A Scoping Review and Crime Script Analysis of Online Romance Scams. *Trauma, Violence, & Abuse*, *0*(0), 1–17. https://doi.org//doi.org/10.1177/15248380251361046

Smart, C. H. (2025). *Untangling Prince Charming : The Role of Scambaiters in the Performance of Romance Fraud and Pig-Butchering Scams*.

Soares, A. B., & Lazarus, S. (2024). Examining fifty cases of convicted online romance fraud offenders. *Criminal Justice Studies*, 37(4), 328–351. https://doi.org/10.1080/1478601X.2024.2429088

Sorell, T. (2019). Scambaiting on the Spectrum of Digilantism. *Criminal Justice Ethics*, *38*(3), 153–175. https://doi.org/10.1080/0731129X.2019.1681132

Soudijn, M. R. J., & Zegers, B. C. H. T. (2012). Cybercrime and virtual offender convergence settings. *Trends in Organized Crime*, *15*(2–3), 111–129. https://doi.org/10.1007/s12117-012-9159-z

Suarez-Tangil, G., Edwards, M., Peersman, C., Stringhini, G., Rashid, A., & Whitty, M. (2019). Automatically Dismantling Online Dating Fraud. *arXiv:1905.12593 [Cs]*. http://arxiv.org/abs/1905.12593

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words : LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Toma, C. L., & Hancock, J. T. (2012). What Lies Beneath : The Linguistic Traces of Deception in Online Dating Profiles. *Journal of Communication*, *62*(1), 78–97. https://doi.org/10.1111/j.1460-2466.2011.01619.x

Tremblay, P. (2010). *Le délinquant idéal. Performance, discipline, solidarité*. (Liber).

Tremblay, P. L. (2004). *Théorie des associations différentielles de Sutherland*. École de criminologie, Université de Montréal.

Wang, F., & Topalli, V. (2024). Understanding Romance Scammers Through the Lens of Their Victims : Qualitative Modeling of Risk and Protective Factors in the Online Context. *American Journal of Criminal Justice*, *49*(1), 145–181. https://doi.org/10.1007/s12103-022-09706-4

Whitty, M. T. (2015). Anatomy of the online dating romance scam. *Secur J*, *28*(4), 443–455. https://doi.org/10.1057/sj.2012.57

Whitty, M. T. (2019). Who can spot an online romance scam? *Journal of Financial Crime*, *26*(2), 623–633. https://doi.org/10.1108JFC06-2018-0053

Whitty, M. T., & Buchanan, T. (2012). The Online Romance Scam : A Serious Cybercrime. *Cyberpsychology, Behavior, and Social Networking*, *15*(3), 181–183. https://doi.org/10.1089/cyber.2011.0352

Zbinden, R., Beaudet-Labrecque, O., Grandjean, F., Gobeil, C., Brunoni, L., Décary-Hétu, D., & Cretu-Adatte, C. (2023). *Scambaiting as a Preventive Tool in the Fight against Cyberfrauds : The Case of Romance Scams*. *8*(1).

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications. *Group Decision and Negotiation*, *13*(1), 81–106. https://doi.org/10.1023/B:GRUP.0000011944.62889.6f

Zingerle, A., & Kronman, L. (Éds.). (2018). Internet Crime and AntiFraud Activism : A Hands-On Approach. In M. Gupta (Ed.), *Security and Privacy Management, Techniques, and Protocols* (p. 322–336). IGI Global. https://doi.org/10.4018/978-1-5225-5583-4