Protocole d'annotation multi-label pour une nouvelle approche à la génération de réponse socio-émotionnelle orientée-tâche

Lorraine Vanel [1, 2], Alya Yacoubi [2] Chloé Clavel [1]
(1) LTCI, Telecom-Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
(2) Zaion, 75008 Paris, France
lorraine.vanel@telecom-paris.fr, chloé.clave@telecom-paris.fr,
ayacoubi@zaion.ai

RÉSUMÉ

Depuis l'apparition des systèmes conversationnels, la modélisation des comportements humains constitue un axe de recherche majeur afin de renforcer l'expression des attributs émotionnels de ces systèmes. En nous intéressant aux agents conversationnels génératifs orientés-tâches, nous proposons une nouvelle approche pour rendre la réponse générée plus pertinente au contexte émotionnel de l'interlocuteur. Cette approche consiste à ajouter une étape supplémentaire de prédiction de labels pour conditionner la réponse générée et assurer sa pertinence au contexte socio-émotionnel de l'utilisateur. Nous proposons une formulation de cette nouvelle tâche de prédiction en nous appuyant sur un protocole d'annotation de données que nous avons conçu et implémenté. À travers cet article, nous apportons les contributions suivantes : la formulation de la tâche de prédiction de labels socio-émotionnels et la description du protocole d'annotation associé. Avec cette méthodologie, nous visons à développer des systèmes conversationnels socialement pertinents et indépendants.

ABSTRACT

Multi-label annotation protocol for a new approach to task-oriented socio-emotional response generation

Ever since the emergence of conversational systems, modeling human behaviors has been a major research topic, aiming to improve the expression of the social and emotional attributes of these systems. With a focus on task-oriented generative conversational agents, we propose a new approach to make the generated response more relevant to the emotional context of the interlocutor. This approach consists in adding an additional label prediction step to condition the generated response and ensure its consistency to the user's socio-emotional context. We propose a formulation of this new prediction task based on a data annotation protocol that we have designed and implemented. In this paper, we introduce the following contributions : the formulation of the socio-emotional label prediction task and the description of the associated annotation protocol. With this methodology we aim at developing socially relevant and independent conversational systems.

MOTS-CLÉS : Génération de Réponse Émotionnelle ; Dialogue Social ; Système Conversationnel Socio-Émotionnel ; Prédiction Socio-Émotionnelle ; Protocole d'Annotation.

KEYWORDS: Emotional Response Generation ; Affective Dialogue ; Social Dialogue ; Socio-Emotionnal Conversational Systems ; Socio-Emotional Label Prediction ; Annotation Protocol.

1 Introduction

L'essor des technologies de traitement du langage a favorisé l'émergence de nombreux assistants virtuels et autres systèmes conversationnels utilisés dans le secteur de la relation client (Ram *et al.*, 2018; Gnewuch *et al.*, 2017), de la formation ou encore de la médecine (Harilal *et al.*, 2020; Adikari *et al.*, 2022). Traditionnellement basés sur des systèmes de règles ou sur une architecture modulaire composée de briques d'intelligence artificielle séparées, une nouvelle ère de l'IA se dessine : celle de la génération automatique de réponse permettant ainsi un dialogue personnalisé et naturel. Cependant, cette approche implique plusieurs défis techniques, tels que la difficulté à combiner l'aspect orienté-tâche avec la dimension émotionnelle des interactions (Clavel *et al.*, 2022). C'est cette prise en compte spontanée de l'émotion latente, naturelle dans un dialogue entre humains, que nous souhaitons apprendre à nos systèmes. Nous faisons donc l'hypothèse qu'utiliser des données conversationnelles humain-humain est l'approche optimale pour entraîner des modèles capables de détecter la réaction émotionnelle de l'interlocuteur et de prendre en considération le contexte social de l'interaction.
Dans la littérature des systèmes de génération de dialogues sociaux et émotionnels, de nombreux travaux proposent de classer les tours de paroles avec des labels comme des émotions, des actes de dialogues ou encore des stratégies de dialogue afin d'assurer la cohérence de la conversation (Li *et al.*, 2017; Welivita *et al.*, 2021). Bien que dans la majorité des cas, un unique label est donné par tour de parole, celui-ci est souvent constitué de plusieurs segments consécutifs qui présentent des stratégies émotionnelles et dialogiques différentes et logiquement dépendantes.
Dans cet article, nous présentons une nouvelle approche pour la tâche de génération de réponse socio-émotionnelle, où une séquence de labels consécutifs est prédite pour plannifier et conditionner la génération. Cela se traduit par l'ajout d'une étape de prédiction de ces labels représentant les comportements attendus dans la réponse du système. Nous proposons les contributions suivantes :

— La formulation de la tâche additionnelle de prédiction de label socio-émotionnel du prochain tour de parole.

— La description d'un protocole d'annotation d'un corpus socio-émotionnel.

Nous commençons par faire une revue des travaux effectués dans le domaine des systèmes de dialogues socio-émotionnels. Ensuite, nous formalisons la tâche de prédiction de labels socio-émotionnels pour détailler le protocole d'annotation adapté, avant d'analyser un corpus annoté selon ces directives.

2 État de l'art

2.1 Les modèles de génération de dialogue émotionnel dans la littérature

Les modèles neuronaux appris sur des corpus de grande taille ont permis d'accélérer l'essor du domaine de la génération de dialogue (Shuster *et al.*, 2022; Zhang *et al.*, 2019; Thoppilan *et al.*, 2022) Certains de ces modèles ont été utilisés pour générer des dialogues socio-émotionnels. Dans ce cas, une annotation émotionnelle des données d'apprentissage est souvent nécessaire. Rashkin *et al.* (2018) proposent un *dataset* annoté en émotions au niveau de la conversation et évalué sur un modèle Transformers. Ce label émotionnel couvre donc la conversation entière, et malgré la précision de la liste de labels utilisée, cette approche ne permet pas d'observer les comportements émotionnels au niveau du tour de parole. Kumar *et al.* (2018) montrent que l'utilisation des actes de dialogue en tant que labels augmente considérablement les performances des systèmes conversationnels. Cependant,

la dimension émotionnelle, au cœur du dialogue humain, n'est pas représentée par ces actes de dialogue. C'est ainsi que des jeux de données comme DailyDialog (Li *et al.*, 2017) (utilisé par Zandie & Mahoor (2020) dans leur modèle EmpTransfo, par exemple) ou le Emotional Dialogue in OpenSubtitles (EDOS) par Welivita *et al.* (2021), présentant une double annotation des stratégies de dialogues et des émotions, capturent les deux aspects que nous cherchons à étudier. Aussi complets soient-ils, ces jeux de données sont composés d'interactions en domaine ouvert et scriptées, ce qui biaise l'authenticité des réponses émotionnelles. De plus, ces corpus ne présentent qu'un unique label par tour de parole, ce qui ne permet pas de représenter les évolutions dynamiques au sein d'un même tour.
Nous avons donc décidé d'explorer les différentes manières d'annoter des données conversationnelles, afin de mettre au point un protocole d'annotation incluant les stratégies de dialogue et les émotions. Cette double annotation nous permet d'exprimer les changements et relations entre les stratégies conversationnelles (stratégies de dialogue et émotions).

2.2 Travaux d'annotation dans la littérature

Certains travaux se sont intéressés à la labellisation des conversations pour conditionner la réponse. Nous énumérons les approches les plus étudiées dans la littérature.

Collecte des données

— ***Crowd-sourcing*** Appliqué à la collecte de données, le *crowd-sourcing* est une méthode participative où un groupe de personnes contribue à la création d'échantillons de données. Les données collectées sont généralement des interactions humains-humain (H-H). Les données sont collectées en faisant interagir deux participants en suivant des directives précises : le locuteur met en place une situation, souvent initiée par un prompt émotionnel (Rashkin *et al.*, 2018; Liu *et al.*, 2021) et l'auditeur doit répondre en conséquence, sans connaître le prompt initial. Les systèmes de dialogue sont formés pour jouer le rôle d'auditeurs.

— **Extraction du Web** Une autre façon courante de collecter des données consiste à extraire des informations de sources en ligne. Dans le cas des données textuelles, il s'agit souvent de messages et de commentaires récupérés sur les réseaux sociaux et il s'agit donc de discours naturel entre humains (Zhong *et al.*, 2020; Mazaré *et al.*, 2018). Elles peuvent également provenir d'autres sources, comme OpenSubtitles (Welivita *et al.*, 2021) où les données sont scénarisées. Les données extraites de ces sites web ne sont généralement pas étiquetées et des processus d'annotation doivent être conçus pour annoter les corpus.

— **Enregistrements de conversations** Cette approche peut être utilisée pour récupérer des archives de conversation humain-humain à partir de données de centres d'appels, comme dans Clavel *et al.* (2013). Ces données sont moins accessibles, car cette pratique nécessite d'avoir les moyens de déployer de tels services ou de demander des données à une entreprise disposant de telles ressources. Même dans ce cas, les données sont généralement confidentielles et ne peuvent donc pas être partagées en tant que jeux de données publics, à moins que le consentement de l'utilisateur ait été donné et que les données aient été correctement anonymisées.

Annotation des données Il existe plusieurs approches de l'annotation des données : elles diffèrent selon le point de vue de l'annotateur, ou encore par les ressources nécessaires à la tâche d'étiquetage.

— **Annotation externe** L'annotation externe, ou du point de vue observateur, implique que le label est donné après analyse par un parti indépendant. Cette approche peut être utilisée sur tout type de données.
  — **Annotation manuelle** Cette approche consiste à entièrement annoter un jeu de données par des experts humains ou des annotateurs qui ont été formés à la tâche spécifique d'annotation. Li *et al.* (2017) présentent un jeu de données de 13K, DailyDialog, qui a été annoté par 3 experts ayant une bonne compréhension de la théorie du dialogue et de la communication, et qui ont été formés aux directives de la tâche particulière (c'est-à-dire l'annotation des émotions et des actes de dialogue). Toutefois, l'annotation manuelle d'un large corpus peut être très coûteuse en temps et en ressources matérielles.
  — **Annotation semi-automatique** Associer l'annotation manuelle à l'usage de modèles permet d'accélérer la tâche d'annotation et d'alléger la charge de travail des juges humains (Lu *et al.*, 2021; Welivita *et al.*, 2021). Il existe de nombreuses manières différentes de réaliser une telle annotation. En général, la première étape consiste à faire annoter par des juges humains une petite fraction des dialogues collectés. Ce sous-ensemble est ensuite utilisé pour entraîner un modèle de classification qui peut soit automatiquement annoter le reste du corpus considéré, ou proposer les labels les plus probables pour chaque exemple du corpus restant. Dans le deuxième cas, c'est aux annotateurs humains de choisir parmi les annotations proposées par le modèle de classification, pour apporter la décision finale.
— **Annotation interne** L'annotation est dite interne lorsque le label est directement dérivé de la source de la donnée.
  — ***Crowd-sourcing*** C'est la principale méthode d'annotation pour ces données, où les émotions et les étiquettes de stratégies de dialogue associées aux données peuvent être directement dérivées des instructions données aux annotateurs (Rashkin *et al.*, 2018; Liu *et al.*, 2021). De plus, Liu *et al.* (2021) recueillent les réponses aux enquêtes soumises aux participants pendant le processus de collecte, tant du côté de l'auditeur que du locuteur. Cela permet de recueillir davantage de données telles que la notation de l'empathie et les stratégies de dialogue au niveau de l'énoncé.
  — **Données extraites d'internet** Zhong *et al.* (2020) utilisent le contexte dans lequel les données web ont été postées et extraient les messages et les commentaires sur deux subreddits : `/r/happy` et `/r/offmychest`. L'environnement original de Reddit fournit donc une étiquette et ce qu'il reste à faire est un contrôle de qualité en demandant à des annotateurs humains d'annoter un petit ensemble de conversations : 100 du Reddit `/r/happy`, 100 de `/r/offmychest` et pour le contrôle, 100 de `/r/casualconversations`.
  — **Enregistrements de conversation** Dans ce cas, l'annotation peut venir d'un retour de l'utilisateur. En effet, certains bots déployés demandent un retour sur la satisfaction des clients, soit directement, soit par le biais de sondages. Ces informations peuvent être utilisées pour annoter certaines conversations (Maslowski *et al.*, 2017; Guibon *et al.*, 2021).

La différence de point de vue de l'annotateur est particulièrement marquée lorsque les labels considérés sont aussi subjectifs que l'émotion. L'annotateur interne peut étiqueter avec précision l'émotion qu'il ressent et exprime dans les données, même si cela vient souvent au prix d'interactions jouées et scénarisées. Un annotateur externe fait une hypothèse quant à l'émotion exprimée, et sa perception est colorée par ses expériences culturelles et ses sensibilités personnelles. Cependant, cette approche est accessible et peut être réalisée sur de nombreux formats de données, ce qui laisse la possibilité d'annoter, à posteriori, des données humain-humain spontanées. Il faut donc garder à l'esprit que

les deux approches comportent leurs propre biais, que ce soit au niveau de la nature scriptée des interaction ou du biais de l'annotation en elle-même. Cette subjectivité de la tâche impacte également les attentes au niveau des scores inter-annotateurs attendus.

2.3 Stratégies et Labels Socio-Émotionnels

Nous présentons enfin les différentes stratégies utilisées dans les approches de génération conversationnelle socio-émotionnelle, ainsi que les différentes manières dont celles-ci sont représentées en tant que labels dans les différents jeux de données disponible dans la littérature. Les jeux de données que nous citons sont en anglais, car nous n'avons trouvé que peu de ressources conversationnelles et annotées en labels sociaux ou émotionnels en français. Nous rappelons que dans cette étude, nous définissons la notion de dialogue "social" par les aspects relationnels liés aux différentes attitudes sociale et à la communication inter-personnelle.

Stratégies basées sur les émotions

— **Définition** Les stratégies basées sur les émotions font référence aux approches qui relèvent de la détection, du traitement et de l'expression d'une émotion en réponse à une situation émotionnelle de l'utilisateur. L'une des approches émotionnelles les plus représentées dans la littérature est l'usage de l'empathie (Fung *et al.*, 2018; Wang *et al.*, 2021; Hosseini & Caragea, 2021), définie comme la capacité à se mettre à la place de son interlocuteur et de percevoir ce qu'iel ressent (Cuff *et al.*, 2016).

— **Labels** Il existe plusieurs manières d'annoter les émotions. Le premier niveau est le sentiment, en annotant la polarité positive ou négative (Lu *et al.*, 2021). Pour ce qui est des émotions plus fines, de nombreuses études font référence à diverses théories de la psychologie pour la classification des émotions, mais il n'y a pas de consensus sur la manière de définir et de classer les émotions dans l'analyse des conversations (Clavel & Callejas, 2016). Li *et al.* (2017) basent leurs annotations sur le modèle d'Ekman (Ekman, 1999), et Liu *et al.* (2021); Rashkin *et al.* (2018) utilisent des théories classiques dérivées des réponses biologiques (Ekman, 1999; Plutchik, 1984) ainsi que des études concentrées sur un ensemble plus large d'émotions subtiles et dépendantes du contexte (Skerry *et al.*, 2015) atteignant jusqu'à 32 étiquettes d'émotions. (Feng *et al.*, 2021) utilisent une classification adaptée aux contextes orientés-tâche.

Stratégies de dialogue

— **Définition** Les stratégies de dialogues sont un ensemble d'actions conversationnelles qui permettent d'exprimer une intention conversationnelle (Galescu *et al.*, 2018; Santos Teixeira & Dragoni, 2022; Liu *et al.*, 2021). Certaines stratégies de dialogue telles qu'*informer* ou *questionner*, se rapprochent des actes de dialogues qui peuvent être interprétés comme la réalisation de ces stratégies. D'autres sont plus émotives, comme *sympathiser* ou *encourager* (Welivita & Pu, 2020).

— **Labels** Certains jeux de données sont doublement annotés en émotions et en stratégies de dialogue. Les deux principales approches que nous avons vues pour ces stratégies sont les stratégies de dialogues elles-mêmes (Hardy *et al.*, 2021; Welivita & Pu, 2020; Liu *et al.*, 2021). Elles peuvent également être associées aux actes de dialogue, résumés et classifiés dans les travaux de Bunt (2006), comme dans le *dataset* DailyDialog (Li *et al.*, 2017).

Stratégies de conception de persona

— **Définition** Une persona est une personnalité fictive dotée de caractéristiques sociales, comme des traits de personnalité ou des préférences (Li *et al.*, 2016). Associer une persona à un système conversationnel revient alors à conditionner les réponses de l'agent pour prendre en compte ces attributs, ce qui permet d'unifier et d'améliorer la cohérence du comportement du système. Plusieurs études relatent les différentes stratégies de conception de ces personas, et de l'influence de certaines caractéristiques sur l'accueil et l'acceptabilité des systèmes chez les utilisateurs (Pradhan & Lazar, 2021; Kim *et al.*, 2019).

— **Labels** Les corpus définissent généralement une persona comme un ensemble de phrases, représentant la personnalité choisie, sur laquelle vont se baser la formulation et le comportement général de l'agent (Mazaré *et al.*, 2018; Zhang *et al.*, 2018). Dans le cadre de données conversationnelles, ces phrases sont souvent collectées à partir de profils d'utilisateurs sur internet, un utilisateur représentant une persona (Zhong *et al.*, 2020; Mazaré *et al.*, 2018).

Ainsi, de nombreuses études s'intéressent à la tâche de génération neuronale de réponse sociale et émotionnelle, en basant leurs approches respectives sur des corpus adéquats. La manière dont ces données sont collectées et annotées en émotions et en stratégies de dialogue est cruciale à la réalisation de la tâche. Le plus souvent, l'annotation est à l'échelle du tour de parole, mais nous souhaitons viser une unité plus fine, en segmentant le tour de parole en segments. Nous nous inspirons des travaux cités pour proposer notre méthodologie d'annotation qui répond au défi technologique que nous souhaitons adresser : la prise en compte dynamique de l'évolution des états émotionnels et dialogiques au cours d'une conversation orientée-tâche.

3 Notre proposition d'annotation multi-label des conversations pour un système de dialogue génératif socio-émotionnel

3.1 Formalisation de la tâche d'annotation en séquence de labels socio-émotionnel

Soit une conversation $C = (c_i)_{i\in[0,t]}$, c_t le tour de parole en cours. Soit SE la liste des labels socio-émotionnels considérés. $\forall i \in [0,t], \exists y_i = (y_i^j)_{j\in[0,l_i]}$, une séquence ordonnée de labels socio-émotionnels associés au tour de parole c_i et $l_i \in \mathbb{N}$ le nombre de labels associés à c_i. $\forall i \in [0,t], \forall j \in [0,l_i], y_i^j \in SE$. Nous considérons maintenant la tâche de prédiction de la séquence y_{t+1} des labels socio-émotionnels associés au tour de parole $t+1$, c_{t+1}. En d'autre termes, il s'agit de prédire l'ensemble :

$$y_{t+1} = (y_{t+1}^j)_{j\in[0,l_{t+1}]} \in SE^{l_{t+1}}$$

Cette séquence sera ensuite utilisée afin d'influencer le système, en prenant en compte les comportement désirés représentés par ces labels pour les générer dans le tour suivant.

3.2 Conception du protocole d'annotation

Dans le but de développer un système pour accomplir la tâche formulée ci-dessus, nous avons conçu un protocole spécifique pour guider la construction d'un corpus annoté avec ces labels socio-émotionnels. Tout d'abord, nous avons recueilli nos données à partir d'enregistrements de conversations entre clients et agents, extraites de nos systèmes déployés dans le cadre industriel. Notre objectif est donc d'annoter manuellement ces conversations avec des labels émotionnels et des stratégies de dialogue. Ce corpus sera utilisé par la suite pour entraîner notre modèle de génération de réponses socialement et émotionnellement pertinentes. Notre jeu de données est composé de 72 conversations téléphoniques en français qui ont été transcrites manuellement par cinq experts en analyse linguistique. Le caractère personnel de ces informations, recueillies en accord avec les régulations RGPD et le consentement des clients, ne nous permets pas de partager ce *dataset*. Cette même équipe d'analystes a également mené la tâche d'annotation de ce corpus textuel.

3.2.1 Préparation des données

Après une revue approfondie de la littérature des systèmes de dialogues socio-émotionnels présentée dans nos travaux précédents (Vanel. *et al.*, 2023), nous avons composé une première liste d'étiquettes pour les émotions et les stratégies de dialogue. Nous avons ensuite analysé nos données pour sélectionner les étiquettes pertinentes. Pour ce faire, nous avons annoté un échantillon de conversations avec la liste complète pour noter les labels manquants ou superflus. Par exemple, dans cette étape, nous avons ajouté la stratégie de politesse, pour indiquer les formules de politesse comme les salutations et les remerciements. Après cette étape, nous avons pu établir la liste finale SE des labels socio-émotionnels. Ces labels ont été annotés à deux niveaux : au niveau du tour de parole (stratégies de dialogue et émotions) et au niveau de la conversation (indices de satisfaction globale). Nous décrivons ces tâches d'annotation plus en détail ci-dessous.

3.2.2 Tâches d'étiquetage

Au niveau du tour de parole Comme indiqué sur la Figure 1, nous cherchons à annoter une liste de labels, issue de la fusion des annotations issues des deux tâches suivantes :

— **Émotions** Pour cette tâche, nous demandons aux analystes d'étiqueter les émotions exprimées à chaque tour de conversation. Les experts lisent chaque tour de conversation, en analysant leur contenu sémantique. Si une émotion est détectée chez un locuteur, l'annotateur note l'émotion et le segment textuel porteur de l'émotion. À la fin de cette tâche, toutes les conversations sont étiquetées en émotions et leurs "indices sémantiques", au niveau du tour de parole.

— **Stratégies de dialogue** Pour les stratégies de dialogue, nous divisons l'annotation en deux tâches : l'annotation des tours de l'agent et l'annotation des tours du client. Ces tâches sont similaires, mais elles sont conduites séparément. Chaque stratégie de dialogue a un code, par exemple la stratégie "Information" est codée par la lettre "I". Un analyste passe sur chaque tour de l'interlocuteur sélectionné, et annote chaque stratégie en les délimitant par des balises. Un tour de parole est annoté dans son entièreté, avec une ou plusieurs stratégies de dialogue consécutives. Cette annotation permet de diviser un tour de parole en une séquence de segments consécutifs et sans chevauchements.

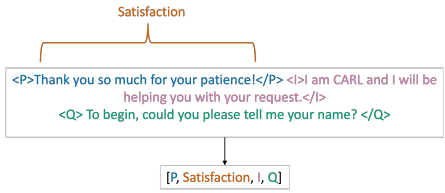

FIGURE 1 – Annotation d'un tour de parole en émotion et stratégie de dialogue.
Ici, P code les formules de politesse, I l'information et Q la question.

Au niveau de la conversation Pour l'annotation au niveau de la conversation, nous avons défini trois aspects qui permettent d'évaluer le ressenti global de la conversation sur plusieurs plans. Ces indices prennent en considération le point de vue du client et la qualité du travail de l'agent. Ils permettent également le filtrage des conversations, car nous souhaitons entraîner nos modèles sur de bons exemples de conversations avec un support agent de bonne qualité pour assurer la satisfaction du client au terme de l'interaction.

— **Satisfaction de l'utilisateur** (Satisfait - Neutre - Insatisfait) annotée par l'annotateur de la stratégie de dialogue du client, à la fin de l'annotation du fichier. Cet indicateur se base sur le comportement des clients et de leur réaction à l'interaction globale avec l'agent.

— **Qualité du support de l'agent** (Bon - Neutre - Mauvais) annotée par l'annotateur de la stratégie de dialogue de l'agent. Cet indicateur mesure le comportement de l'agent, et la justesse de ses réponses et interactions avec le client selon la situation de ce dernier.

— **Statut de résolution du problème** (Résolu - Incertain - Non résolu) annoté par l'annotateur de la stratégie de dialogue du client. Ce dernier indicateur signale l'état final de la demande du client, si le motif d'appel a été identifié et résolu par l'agent.

3.2.3 Réalisation de l'annotation

Phase de Calibrage

— **Calibrage** L'objectif de cette phase est d'initier les annotateurs à la tâche d'annotation. Les cinq annotateurs évaluent chacun l'échantillon sélectionné pour les trois tâches : émotions, stratégies de dialogue et indices de satisfaction globale. Ces fichiers ont été échantillonnés aléatoirement, tout en essayant de garder une bonne distribution des longueurs des conversations.

— **Mesure d'accord** Nous avons calculé l'alpha de Krippendorff pour évaluer les niveaux d'accord et sommes arrivés à un alpha de 0,674. La raison pour laquelle nous avons choisi l'alpha de Krippendorff, plutôt que le Kappa de Cohen par exemple, est car il s'agit d'une métrique qui prend en compte l'annotation multi-label et les annotateurs multiples. Cependant, elle ne tient pas compte de l'ordre des labels, uniquement du compte global de chaque annotation par tour.

Annotation complète Après avoir confirmé que les mesures d'accord étaient supérieures à notre seuil d'acceptation, nous avons lancé le processus d'annotation sur les échantillons restants.

3.3 Analyse du Corpus Annoté

Le corpus Le jeu de données annoté est composé de 67 conversations, ce qui représente 3051 tours de parole et 109 841 tokens. L'annotation émotionnelle a été faite avec 20 labels émotionnels secondaires, mais pour des raisons de lisibilité et de clarté, nous avons décidé de mener cette analyse en reprenant les émotions primaires correspondantes (par exemple *agacement* est une émotion secondaire de *peur*). Notre liste de labels est donc composée de 6 labels émotions et de 15 stratégies de dialogue, présentés en Table 3. La longueur moyenne des tours de parole est de 29 tokens. Ces tours de parole sont divisés en segments délimités par ces stratégies de dialogue consécutives, et leur longueur moyenne est de 17 tokens. Le nombre de labels socio-émotionnels l_i moyen par tour de parole est de 2.5 avec un l_i maximum de 23 labels socio-émotionnels, et un l_i minimum de 1 (chaque tour porte au moins un label de stratégie de dialogue). Nous donnons plus de détails sur les différences entre agent et client dans la Table 1. Nous remarquons également que les agents ont tendance à produire des tours de paroles plus longs et riches en stratégies. Une médiane de 2 labels par tour de parole valide notre approche multi-label pour la planification du prochain tour de parole. En effet, il y a en général plusieurs stratégies qui se succèdent, et reproduire le comportement d'un agent implique la prise en considération de ces multiples éléments successifs au sein d'un même tour de parole.

Les 3051 tours de paroles sont donc segmentés en 6852 segments dénotés par les stratégies de dialogue, auxquels se superposent 785 labels émotionnels. Cette différence entre le nombre d'émotions et de stratégies de dialogue s'explique par le protocole : plusieurs stratégies de dialogues sont données par tour de parole et tous les tours de parole sont entièrement annotés en stratégie de dialogue (il n'existe aucun tour de parole neutre sans aucune annotation de stratégie), alors que tous les tours ne sont pas porteurs d'émotion. Lorsqu'ils le sont c'est en général une unique émotion qui est représentée. Nous notons que la majorité des conversations propose un service qui convient au client et que dans 63 conversations, l'agent apporte un bon niveau de support.

Distribution des labels Nous avons ensuite étudié la distribution des labels annotés. La liste complète de ces labels est présentée dans la Table 3. Pour les émotions, la satisfaction est l'émotion dominante dans le discours des agents alors que pour les clients, c'est l'agacement. Certaines émotions sont très peu observées, comme la rage ou la détresse, ce qui est attendu dans le contexte des interactions que nous avons choisi. Pour ce qui est des stratégies de dialogue, la stratégie *information* est la plus représentée chez les deux interlocuteurs. En effet, celle-ci est utilisée pour apporter une information, répondre à une question ou indiquer un choix. Au niveau de la fréquence d'utilisation des stratégies,

	Agent	Client
Minimum	1	1
Maximum	16	23
Moyenne	2.7	2.3
Médiane	2	2

TABLE 1 – Statistiques liées à l_i, le nombre de labels socio-émotionnels associés au tour de parole i.

Stratégie	Total	Agent	Client
Émotions	785	382	403
Stratégies de Dialogue	6852	3582	3270
Total	7637	3964	3673
Indice	**Satisfait**	**Neutre**	**Insatisfait**
Satisfaction Utilisateur	45	15	7
	Bon	**Neutre**	**Mauvais**
Qualité Support Agent	64	3	0
	Résolu	**Incertain**	**Non Résolu**
Résolution du Problème	36	28	3

TABLE 2 – Nombre de labels annotés au niveau du tour de parole et au niveau de la conversation.

(a) Émotions

Emotion	Total	Agent	Client
Colère (*impatience, agacement, rage*)	203	42	161
Surprise	32	16	16
Peur (*stress, inquiétude, confusion*)	115	26	89
Joie (*soulagement, satisfaction*)	381	271	110
Tristesse (*déception, résignation*)	45	19	26
Dégoût (*mépris, moquerie*)	9	5	4

(b) Stratégies de Dialogue

Stratégie (Code)	Total	Agent	Client
Accord (A)	20	7	13
Aggressivité (AGR)	11	0	11
Back-Channeling (BC)	645	286	359
Correction (C)	66	18	48
Désaccord (D)	8	1	7
Encouragement (E)	52	43	9
Hors-Sujet (HS)	235	92	143
Information (I)	3023	1368	1655
Politesse (P)	721	367	354
Proposition de Suggestions (PS)	86	62	24
Question (Q)	1213	906	307
Reformulation (R)	439	328	111
Sympathie (S)	49	42	7
Self-Disclosure (SD)	240	55	185
Autre (U)	44	7	37

TABLE 3 – Compte de labels par émotion et stratégie de dialogue.

nous remarquons que pour le client ce sont celles de *question* et d'*information* qui dominent. Cela correspond au script général que les agents suivent pour qualifier puis traiter la demande client en réponse à une sollicitation client. Pour les clients, c'est la stratégie d'*information* qui est majoritaire, et constitue souvent une réponse aux questions de l'agent qui les guide à travers le processus. En effet, les 906 questions de l'agent sont réparties sur 773 tours de paroles différents (un unique tour de parole peut porter plusieurs questions), dont 619 sont suivis par une information donnée par le client.

Patterns dans les Labels Nous avons observé des patterns récurrents dans la succession des labels dans un tour de parole. Nous nous intéressons dans un premier temmps aux successions de 2 et 3 labels répétés dans le même ordre dans les énoncés du corpus. Les motifs les plus communs sont présentés en Table 4.

(a) k = 2

Pattern	Count
Information, Question	222
Information, Information	178
Question, Information	168
Back-Channeling, Information	133
Back-Channeling, Back-Channeling	127
Information, Joie	126
Reformulation, Information	100
Information, Politeness	98
Politesse, Information	85
Information, Back-Channeling	85

(b) k = 3

Pattern	Count
Information, Question, Information	92
Question, Information, Question	51
Back-channeling, Back-channeling, Back-channeling	34
Back-channeling, Back-channeling, Information	31
Information, Joie, Question	24

TABLE 4 – Motifs les plus communs pour 2 labels et 3 labels successifs dans le même tour de parole.

La plupart de ces motifs inclue des stratégies de dialogue, telles qu'*Information* ou *Back-Channeling*. Cependant, nous remarquons que les émotions sont aussi représentées, notamment par le label *Joie*. Les labels ont été annoté séparément, ce qui signifie qu'une succession de deux questions sera annoté 'Question, Question', ce qui explique pourquoi les labels peuvent se répété au sein d'un même tour (par exemple, *Back-Channeling* est une stratégie répétitive par nature, car cela implique d'interjecter pendant une prise de parole d'un interlocuteur pour le montrer qu'on prête attention). Ces patterns

nous donnent une idée de comment la plupart des réponses sont planifiées et exécutées de manière consciente ou non par les agents humains.

4 Discussion

Nous nous sommes posé la question du format des données d'entrée et de sortie de cette tâche de prédiction, particulièrement au niveau de l'unité d'annotation considérée. Dans cet article, nous utilisons le tour de parole, mais notre protocole d'annotation permet de considérer une autre unité : le segment. Chaque tour de parole est annoté en une ou plusieurs stratégies de dialogues consécutives, qui délimitent des segments. Cela impliquerait donc la définition d'un ensemble de segments c_i^j associés au tour de parole c_i, et que, pour chaque segment j, il existe une liste de labels $y_i^{j,k}$ de labels socio-émotionnels associés à ce segment. La tâche de génération serait donc reportée à l'échelle du segment, ce qui ajoute la difficulté supplémentaire de fusionner ces différents segments pour former l'unique énoncé final c_{t+1}. Nous pensons que la planification de la génération du prochain tour de parole dans son ensemble est plus pertinente, et nous avons donc choisi de conserver l'échelle du tour de parole, sans segmentation.
Il est important de noter que l'annotation comprend des biais liés aux expériences personnelles aussi bien que culturelles des annotateurs, qui peuvent influencer leurs perception des émotions et des interactions. Nous avons mis au point la phase de calibrage pour pouvoir privilégier au mieux la communication, l'extraction et l'analyse des possibles divergences, biais et incompréhensions. Cette première annotation nous a permis, au-delà de sortir les mesures d'accord, d'établir des règles et des consignes plus précises d'annotation, agrémentées d'exemples issus des données. De plus, nous avons choisi une fine granularité des annotations, pour améliorer l'explicabilité du contenu social généré. Enfin, nous n'avons pas encore ajouté de profils persona, mais nous considérons explorer cette piste à l'avenir. Nous souhaitons explorer certaines pistes, comme identifier les différents agents humains dans nos données et baser des personas à partir des données associées à ces agents. Nous pouvons également concevoir des fiches de personas fictives élaborées en collaboration avec les clients. Il serait ainsi intéressant d'étudier les différences dans les stratégies de dialogues et labels émotionnels observés selon les personas implémentées.

5 Conclusion et prochains travaux

Dans cet article, nous présentons une nouvelle approche pour la tâche de génération, en ajoutant une étape de prédiction d'une séquence de labels d'émotion et de stratégie de dialogue attendus dans le tour suivant. Nous détaillons également le protocole d'annotation associé que nous avons implémenté sur un corpus orienté-tâche. À l'issu de l'analyse de ce jeu de données annoté, la richesse du contenu émotionnel et dialogique nous conforte dans notre choix de données humain-humain, spontanées et issues d'un contexte orienté-tâche réel. L'analyse montre également qu'un tour de parole est souvent porteur de plus d'un seul label socio-émotionnel (médiane de 2 labels par tour) et l'importance des liens entre ces différents labels. Cela justifie notre approche de prédiction d'une séquence multi-label pour planifier et générer le prochain tour de parole du système conversationnel.
Nos futurs travaux utiliseront ce jeu de données annoté pour développer un système conversationnel génératif orienté-tâche capable de s'adapter de manière dynamique aux évolutions du contexte émotionnel et social de la conversation.

Références

ADIKARI A., DE SILVA D., MORALIYAGE H., ALAHAKOON D., WONG J., GANCARZ M., CHACKOCHAN S., PARK B., HEO R. & LEUNG Y. (2022). Empathic conversational agents for real-time monitoring and co-facilitation of patient-centered healthcare. *Future Generation Computer Systems*, **126**, 318–329. DOI : https ://doi.org/10.1016/j.future.2021.08.015.

BUNT H. (2006). Dimensions in dialogue act annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy : European Language Resources Association (ELRA).

CLAVEL C., ADDA G., CAILLIAU F., GARNIER-RIZET M., CAVET A., CHAPUIS G., COURCINOUS S., DANESI C., DAQUO A.-L., DELDOSSI M., GUILLEMIN-LANNE S., SEIZOU M. & SUIGNARD P. (2013). Spontaneous speech and opinion detection : mining call-centre transcripts. *Language Resources and Evaluation*, **47**(4), 1089–1125.

CLAVEL C. & CALLEJAS Z. (2016). Sentiment analysis : From opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing*, **7**(1), 74–93. DOI : 10.1109/TAFFC.2015.2444846.

CLAVEL C., LABEAU M. & CASSELL J. (2022). Socio-conversational systems : Three challenges at the crossroads of fields. *Frontiers in Robotics and AI*, **9**. DOI : 10.3389/frobt.2022.937825.

CUFF B., BROWN S., TAYLOR L. & HOWAT D. (2016). Empathy : A review of the concept. *Emotion Review*, **8**, 144–153. DOI : 10.1177/1754073914558466.

EKMAN P. (1999). Basic emotions. *Handbook of cognition and emotion*, **98**(45-60), 16.

FENG S., LUBIS N., GEISHAUSER C., LIN H.-C., HECK M., VAN NIEKERK C. & GAŠIĆ M. (2021). Emowoz : A large-scale corpus and labelling scheme for emotion recognition in task-oriented dialogue systems. DOI : 10.48550/ARXIV.2109.04919.

FUNG P., BERTERO D., XU P., PARK J. H., WU C.-S. & MADOTTO A. (2018). Empathetic dialog systems. In *The international conference on language resources and evaluation. European Language Resources Association*.

GALESCU L., TENG C. M., ALLEN J. & PERERA I. (2018). Cogent : A generic dialogue system shell based on a collaborative problem solving model. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, p. 400–409.

GNEWUCH U., MORANA S. & MAEDCHE A. (2017). Towards designing cooperative and social conversational agents for customer service. In *ICIS*.

GUIBON G., LABEAU M., FLAMEIN H., LEFEUVRE L. & CLAVEL C. (2021). Few-shot emotion recognition in conversation with sequential prototypical networks. *CoRR*, **abs/2109.09366**.

HARDY A., PARANJAPE A. & MANNING C. D. (2021). Effective social chatbot strategies for increasing user initiative. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, p. 99–110.

HARILAL N., SHAH R., SHARMA S. & BHUTANI V. (2020). Caro : An empathetic health conversational chatbot for people with major depression. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, p. 349–350, New York, NY, USA : Association for Computing Machinery. DOI : 10.1145/3371158.3371220.

HOSSEINI M. & CARAGEA C. (2021). It takes two to empathize : One to seek and one to provide. In *Proceedings of the AAAI Conference on Artificial Intelligence. To appear*.

KIM H., KOH D. Y., LEE G., PARK J.-M. & LIM Y.-K. (2019). Designing personalities of conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, p. 1–6, New York, NY, USA : Association for Computing Machinery. DOI : 10.1145/3290607.3312887.

KUMAR H., AGARWAL A. & JOSHI S. (2018). Dialogue-act-driven conversation model : An experimental study. In *Proceedings of the 27th International Conference on Computational Linguistics*, p. 1246–1256, Santa Fe, New Mexico, USA : Association for Computational Linguistics.

LI J., GALLEY M., BROCKETT C., SPITHOURAKIS G. P., GAO J. & DOLAN B. (2016). A persona-based neural conversation model. *arXiv preprint arXiv :1603.06155*.

LI Y., SU H., SHEN X., LI W., CAO Z. & NIU S. (2017). Dailydialog : A manually labelled multi-turn dialogue dataset. DOI : 10.48550/ARXIV.1710.03957.

LIU S., ZHENG C., DEMASI O., SABOUR S., LI Y., YU Z., JIANG Y. & HUANG M. (2021). Towards emotional support dialog systems. *ArXiv*, **abs/2106.01144**.

LU X., TIAN Y., ZHAO Y. & QIN B. (2021). Retrieve, discriminate and rewrite : A simple and effective framework for obtaining affective response in retrieval-based chatbots. In *Findings of the Association for Computational Linguistics : EMNLP 2021*, p. 1956–1969.

MASLOWSKI I., LAGARDE D. & CLAVEL C. (2017). In-the-wild chatbot corpus : from opinion analysis to interaction problem detection. In *ICNLSSP 2017*, p. 115–120, Casablanca, Morocco : ISGA, Institut Supérieur d'inGénierie et des Affaires. HAL : hal-02288505.

MAZARÉ P.-E., HUMEAU S., RAISON M. & BORDES A. (2018). Training millions of personalized dialogue agents. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 2775–2779, Brussels, Belgium : Association for Computational Linguistics. DOI : 10.18653/v1/D18-1298.

PLUTCHIK R. (1984). Emotions : a general psychoevolutionary theory.

PRADHAN A. & LAZAR A. (2021). Hey google, do you have a personality ? designing personality and personas for conversational agents. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI '21, New York, NY, USA : Association for Computing Machinery. DOI : 10.1145/3469595.3469607.

RAM A., PRASAD R., KHATRI C., VENKATESH A., GABRIEL R., LIU Q., NUNN J., HEDAYATNIA B., CHENG M., NAGAR A., KING E., BLAND K., WARTICK A., PAN Y., SONG H., JAYADEVAN S., HWANG G. & PETTIGRUE A. (2018). Conversational ai : The science behind the alexa prize. DOI : 10.48550/ARXIV.1801.03604.

RASHKIN H., SMITH E. M., LI M. & BOUREAU Y.-L. (2018). Towards empathetic open-domain conversation models : A new benchmark and dataset. *arXiv preprint arXiv :1811.00207*.

SANTOS TEIXEIRA M. & DRAGONI M. (2022). A review of plan-based approaches for dialogue management. *Cognitive Computation*, p. 1–20.

SHUSTER K., XU J., KOMEILI M., JU D., SMITH E. M., ROLLER S., UNG M., CHEN M., ARORA K., LANE J., BEHROOZ M., NGAN W., POFF S., GOYAL N., SZLAM A., BOUREAU Y.-L., KAMBADUR M. & WESTON J. (2022). Blenderbot 3 : a deployed conversational agent that continually learns to responsibly engage. DOI : 10.48550/ARXIV.2208.03188.

SKERRY A. E., SAXE R., SKERRY A. E. & SAXE R. (2015). Neural representations of emotion are organized around abstract event features. *Curr. Biol.*

THOPPILAN R., DE FREITAS D., HALL J., SHAZEER N., KULSHRESHTHA A., CHENG H.-T., JIN A., BOS T., BAKER L., DU Y., LI Y., LEE H., ZHENG H. S., GHAFOURI A., MENEGALI M., HUANG Y., KRIKUN M., LEPIKHIN D., QIN J., CHEN D., XU Y., CHEN Z., ROBERTS A., BOSMA M., ZHAO V., ZHOU Y., CHANG C.-C., KRIVOKON I., RUSCH W., PICKETT M., SRINIVASAN P., MAN L., MEIER-HELLSTERN K., MORRIS M. R., DOSHI T., SANTOS R. D., DUKE T., SORAKER J., ZEVENBERGEN B., PRABHAKARAN V., DIAZ M., HUTCHINSON B., OLSON K., MOLINA A., HOFFMAN-JOHN E., LEE J., AROYO L., RAJAKUMAR R., BUTRYNA A., LAMM M., KUZMINA V., FENTON J., COHEN A., BERNSTEIN R., KURZWEIL R., AGUERA-ARCAS B., CUI C., CROAK M., CHI E. & LE Q. (2022). Lamda : Language models for dialog applications. DOI : 10.48550/ARXIV.2201.08239.

VANEL. L., YACOUBI. A. & CLAVEL. C. (2023). A survey of socio-emotional strategies for generation-based conversational agents. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence - Volume 3 : ICAART,*, p. 185–192 : INSTICC SciTePress. DOI : 10.5220/0011632400003393.

WANG Y.-H., HSU J.-H., WU C.-H. & YANG T.-H. (2021). Transformer-based empathetic response generation using dialogue situation and advanced-level definition of empathy. In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, p. 1–5. DOI : 10.1109/ISCSLP49672.2021.9362067.

WELIVITA A. & PU P. (2020). A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, p. 4886–4899, Barcelona, Spain (Online) : International Committee on Computational Linguistics. DOI : 10.18653/v1/2020.coling-main.429.

WELIVITA A., XIE Y. & PU P. (2021). A large-scale dataset for empathetic response generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, p. 1251–1264.

ZANDIE R. & MAHOOR M. H. (2020). Emptransfo : A multi-head transformer architecture for creating empathetic dialog systems. *CoRR*, **abs/2003.02958**.

ZHANG S., DINAN E., URBANEK J., SZLAM A., KIELA D. & WESTON J. (2018). Personalizing dialogue agents : I have a dog, do you have pets too ? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 2204–2213, Melbourne, Australia : Association for Computational Linguistics. DOI : 10.18653/v1/P18-1205.

ZHANG Y., SUN S., GALLEY M., CHEN Y.-C., BROCKETT C., GAO X., GAO J., LIU J. & DOLAN B. (2019). Dialogpt : Large-scale generative pre-training for conversational response generation. DOI : 10.48550/ARXIV.1911.00536.

ZHONG P., ZHANG C., WANG H., LIU Y. & MIAO C. (2020). Towards persona-based empathetic conversational models. *arXiv preprint arXiv :2004.12316*.