



# Un modèle pour la généricité des agents conversationnels vocaux multi-domaines

Maya Medjad, Mathieu Buonomo, Raphaël Szymocha, Frédéric Armetta

## ► To cite this version:

Maya Medjad, Mathieu Buonomo, Raphaël Szymocha, Frédéric Armetta. Un modèle pour la généricité des agents conversationnels vocaux multi-domaines. RJCIA 2023, Jul 2023, Strasbourg, France. <hal-04209690>

**HAL Id: hal-04209690**

**<https://hal.science/hal-04209690v1>**

Submitted on 18 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# Un modèle pour la généricité des agents conversationnels vocaux multi-domaines

Maya Medjad<sup>1 2</sup>, Mathieu Buonomo<sup>2</sup>, Raphaël Szymocha<sup>2</sup>, Frédéric Armetta<sup>1</sup>

<sup>1</sup> Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, F-69622 Villeurbanne, France  
nom.prenom@liris.cnrs.fr

<sup>2</sup> Reecall, reecall.com

5 mai 2023

## Résumé

Les agents conversationnels vocaux sont de plus en plus utilisés pour des services automatisés d'assistance téléphonique (support, réservations, commandes, etc). Après avoir décrit la chaîne de traitement d'un agent conversationnel, nous soulignons des spécificités propres au déploiement de ces systèmes sur plusieurs domaines et proposons un modèle permettant d'étendre les possibilités de paramétrage et de généralisation à différents domaines d'activité.

## Mots-clés

Traitement naturel du langage, Agent conversationnel, Apprentissage profond, Chatbot

## Abstract

Conversational voice agents are increasingly used for automated phone assistance services (support, reservations, orders, etc.). After describing the processing chain of a conversational agent, we highlight specific features for the deployment of such systems over several domains and propose a model to extend the possibilities of parameterisation and generalisation to different domains of activity.

## Keywords

Natural language processing, Conversationnal agent, Deep learning, Chatbot

## 1 Introduction

Les agents conversationnels sont de plus en plus utilisés pour leurs capacités d'interaction en langage naturel avec des utilisateurs, afin de fournir des réponses rapides, pertinentes et personnalisées dans différents domaines (suivi de commandes, etc.). Au-delà de l'automatisation, ceux-ci offrent aussi une meilleure disponibilité lors de fortes sollicitations des services, tout en permettant d'orienter l'utilisateur vers un conseiller lorsque cela est approprié.

Dans le cadre de ce travail, nous nous intéressons au déploiement d'agents conversationnels multi-domaines, qui permettent de traiter des conversations pour des domaines d'application variés. Nous soulignons dans ce travail des spécificités que nous avons identifiées propres au déploiement multi-domaine, et proposons un flux de traitement générique afin de fiabiliser le déploiement de tels systèmes.

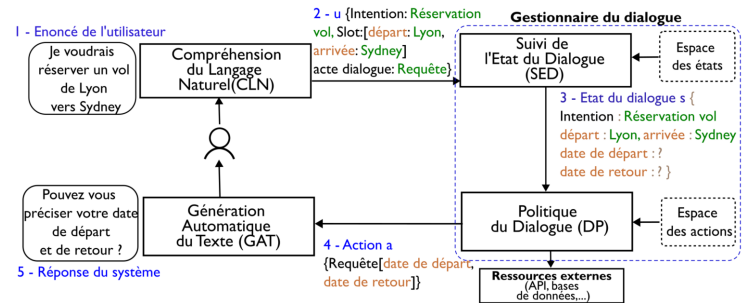


FIGURE 1 – Gestion d'un tour de conversation [1]

En section 2, nous décrivons les méthodes utilisées pour le déploiement de ces systèmes. Nous introduisons ensuite en section 3 notre proposition de modèle pour la mise en application d'un système multi-domaines, et concluons en section 4.

## 2 Les agents conversationnels

La gestion de la conversation s'articule autour de différents composants qui permettent à partir d'un message de l'utilisateur formulé en langage naturel d'identifier l'intention portée par le message, pour le contextualiser au sein de la conversation et fournir une réponse adaptée à l'utilisateur.

Nous retrouvons ces composants sur la figure 1, tel que présenté par [1] :

- Le composant CLN [4] (Compréhension du Langage Naturel) permet d'identifier (souvent grâce à des réseaux de neurones) l'intention de l'utilisateur tout en extrayant des informations utiles à la conversation. Sur la figure, on identifie l'intention de réserver et certaines données utiles telles que la ville de destination.
- L'information ainsi recueillie permet au SED (Suivi de l'État du Dialogue) de contextualiser l'état de progression de la conversation, en identifiant des informations encore non renseignées telle que la date de départ.
- Le DP (Politique du Dialogue) détermine alors la prochaine action du système, tel que la demande d'informations complémentaires permettant

de compléter la transaction.

- Le composant de GAT (Génération automatique de texte) génère quant à lui une réponse appropriée en langage naturel.

Toutes ces étapes peuvent être implémentées de façon modulaire ou fusionnée selon l'approche choisie. Les systèmes de gestion de la conversation peuvent ainsi prendre plusieurs formes, avec des méthodes dites manuelles [6] (la conversation suit alors son cours en parcourant différents états et transitions pré-établies), de l'apprentissage automatique (supervisé ou par renforcement [2]) ou des modèles hybrides [7]. Les méthodes manuelles permettent d'explicitement les étapes de la conversation (données à recueillir, informations à communiquer, etc.). Les méthodes issues du deep learning requièrent beaucoup de données parfois difficiles à rassembler. Dans le domaine des systèmes de conversation orientés but on peut utiliser des approches hybrides pour contrôler le fil de la conversation, avec un paramétrage des étapes de la conversation explicite (diagramme d'états et de transitions) tout en utilisant des classifieurs pour l'identification des intentions de l'utilisateur [9].

Les données auxquelles font face ces systèmes pour identifier les intentions des utilisateurs ne sont jamais connues entièrement à l'avance. On définit les domaines d'application pour lesquels l'agent répondra, ce qui ne correspond pas est alors identifié "hors du champ" [3] (*out of scope*).

Parmi les différentes application propres à ce type de système, le besoin d'étendre le périmètre de compétence de ces agents à plusieurs domaines soulève de nombreux défis [5] tel que la disponibilité des datasets.

Dans cet article, on s'intéresse à étendre la détection des intentions de l'utilisateur à plusieurs domaines, en limitant l'introduction de classifieurs spécifiques difficiles à paramétrer.

### 3 Des agents conversationnels génériques

#### 3.1 La spécificité a un coût

La société Reecall<sup>1</sup> propose différents services à ses clients dans des domaines variés tels que l'immobilier ou le e-commerce. Afin d'adapter les agents conversationnels suivant les besoins spécifiques des clients, la gestion du dialogue s'effectue par une modélisation à états finis des étapes de la conversation, dont les transitions sont déclenchées par les intentions de l'utilisateur identifiées par différents classifieurs (composant CLN).

Un scénario correspond alors à un échange composé de questions et d'actions entre l'agent et son interlocuteur afin de traiter sa demande. De nouveaux domaines sont régulièrement ajoutés au système, ce qui nécessite l'introduction de classifieurs spécifiques, dont l'apprentissage nécessite également différents jeux de données dédiés. La gestion de la conversation doit elle-même être adaptée à ces nouveaux domaines. L'introduction de ces classifieurs peut

J'ai besoin de parler avec un technicien.	Ce serait pour prendre un rendez-vous.	J'ai oublié mon mot de passe.
READ Escalation	CREATE Main Object	UPDATE Account

FIGURE 2 – Détection d'intentions primaire (en bleu) et secondaires (en vert)

être complexe et introduire des anomalies (jeux d'entraînement déséquilibrés, hors sujet (*out-of-scope*) difficiles à définir), la définition de nouvelles politiques de conversation peut également introduire des incertitudes et des délais supplémentaires pour le déploiement. Afin d'améliorer la satisfaction des utilisateurs et de fiabiliser un tel système, nous proposons ainsi un modèle plus générique qui a pour vocation de pouvoir s'appliquer et s'adapter aux différents domaines rencontrés, tout en maintenant des possibilités de spécialisation des services proposés.

#### 3.2 Un modèle multi-domaine plus générique

Nous proposons un modèle de détection d'intentions qui s'appuie sur la détection conjointe d'intentions primaires et secondaires, tel que présenté sur la figure 2. Une interaction entre l'utilisateur et l'agent est ainsi décomposée de cette façon : 1) L'utilisateur émet une demande, 2) l'intention primaire de l'utilisateur est détectée (action à effectuer), 3) l'intention secondaire est détectée à partir de la même phrase formulée (objet concerné), 4) l'agent identifie ainsi le contexte et répond à l'utilisateur suivant la combinaison des intentions primaire et secondaire identifiées.

L'intention primaire et secondaire se complètent donc pour exprimer la demande de l'utilisateur. Les détections des deux classifieurs dédiés permettent de détecter l'entièreté de la demande exprimée par l'utilisateur (par exemple "UPDATE" et "ACCOUNT" sur la figure 2).

Si l'une des deux classifications échoue (ce qui se produit régulièrement en environnement bruité par exemple, l'échec de reconnaissance pouvant être détecté par différentes méthodes de seuils de confiance non détaillées dans cet article, une détection avec un seuil trop faible ne sera pas prise en compte), l'agent conversationnel peut tout de même déclencher un scénario adapté à l'action ou au contexte identifié. Il pourra par exemple demander des précisions sur l'objet de la mise à jour ("UPDATE" détecté), ou inversement demander des précisions sur l'opération à réaliser sur l'objet ("ACCOUNT" détecté).

En procédant ainsi, le classifieur d'intention primaire qui identifie l'action est commun et réutilisable pour tous les clients. Celui associé aux intentions secondaires est malgré tout à différencier suivant le domaine concerné ("*Main Object*" correspondra à un "rendez-vous" pour le domaine des rendez-vous, ou au "produit" pour le domaine du "e-commerce", avec des formulations spécifiques à définir pour l'apprentissage), mais pourront être utilisés pour différents agents conversationnels abordant des thèmes communs, garantissant ainsi une meilleure fiabilité. Le déroulement de la conversation pourra ainsi être défini au cas par cas suivant les combinaisons d'intentions primaires et se-

1. La société Reecall est spécialisée dans la mise en place d'agents conversationnels vocaux [reecall.com](http://reecall.com)

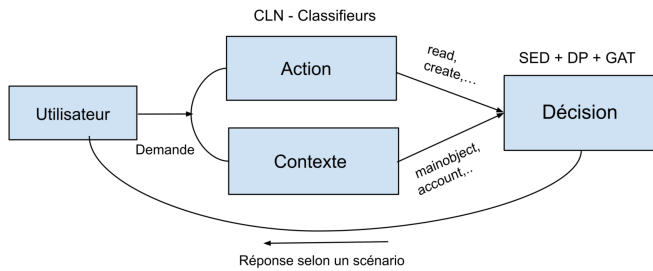


FIGURE 3 – Le système

concordances identifiées détectées au cours de la conversation.

### 3.3 Des actions génériques pour représenter toutes les configurations

Nous avons sélectionné les actions (ou intentions primaires) en s'inspirant du modèle CRUD [8] du domaine des bases de données, dédié à la manipulation d'objets (*create, delete, read, update*). Afin de traiter les actions inconnues, nous complétons cette liste par l'action oos (*out-of-scope*).

Concernant le Contexte (intentions secondaires), nous avons identifié sept classes (*mainobject, invoice, quote, account, payment, escalation, oos*) qui peuvent être explicitées de façon différente suivant le domaine concerné. Ce qui correspond à l'objet principal de l'échange pour un domaine sera qualifié de *mainobject*, *escalation* correspond à la mise en relation avec un opérateur humain, *invoice* concerne la facturation, etc.

Ainsi, si le système détecte "create(ip) account(is)", cela signifie que l'utilisateur souhaite créer un compte, et le scénario suivi permet alors de récupérer les informations complémentaires nécessaires. Le contexte est ainsi identifié par la combinaison de l'intention primaire et de l'intention secondaire (par domaine). Certaines applications nécessitent toutefois l'utilisation d'intentions secondaires issues de plusieurs domaines, les classifieurs dédiés aux intentions secondaires peuvent alors se compléter, et étendre ainsi les contextes qui peuvent être identifiés, et orienter le déroulement de la conversation.

Avec un tel système, nous pouvons donc définir une configuration de base stable et éprouvée, selon les besoins du client pour l'identification des intentions (module CLN), et créer ou adapter la gestion de la conversation (module SED et DP).

## 4 Conclusion

Nous avons présenté dans cet article une mise en application pratique pour le développement d'agents conversationnels vocaux plus génériques. En proposant un système plus générique, adaptable à différents domaines, nous souhaitons fiabiliser le paramétrage et l'agencement des différents classifieurs dédiés à l'identification des intentions de l'utilisateur. Améliorer la qualité de détection des intentions de l'utilisateur est en effet primordial pour garantir la pertinence des réponses apportées à l'utilisateur par le module conversationnel. Cette approche permet également un déploiement plus rapide, tout en gardant la possi-

bilité de personnalisation des scénarios suivant les cas d'utilisation. Le modèle proposé s'intéresse principalement à améliorer le module de reconnaissance des intentions dans un contexte multi-domaine. Pour de futurs travaux, nous pourrions étudier l'extension de la généralisation au module conversationnel, afin d'automatiser le couplage entre la configuration utile à la reconnaissance des intentions, et le module conversationnel, pour les agents conversationnels multi-domaines.

## Références

- [1] Hayet Brabra, Marcos Báez, Boualem Benatallah, Walid Gaaloul, Sara Bouguelia, and Shayan Zamanirad. Dialogue management in conversational systems : a review of approaches, challenges, and opportunities. *IEEE Transactions on Cognitive and Developmental Systems*, 2021.
- [2] Zhi Chen, Lu Chen, Xiang Zhou, and Kai Yu. Deep reinforcement learning for on-line dialogue state tracking. *arXiv preprint arXiv :2009.10321*, 2020.
- [3] L Manik, Zaenal Akbar, Hani Febri Mustika, Ariani Indrawati, Dwi Setyo Rini, Agusdin Dharma Fefirenta, and Tutie Djarwaningsih. Out-of-scope intent detection on a knowledge-based chatbot. *International Journal of Intelligent Engineering and Systems*, 14(5) :446–457, 2021.
- [4] Marjorie McShane. Natural language understanding (nlu, not nlp) in cognitive systems. *AI Magazine*, 38(4) :43–56, 2017.
- [5] Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. Towards scalable multi-domain conversational agents : The schema-guided dialogue dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8689–8696, 2020.
- [6] Cameron Smith, Nigel Crook, Daniel Charlton, Johan Boye, Raul Santos De La Camara, Markku Turunen, David Benyon, Björn Gambäck, Oli Mival, Nick Webb, et al. Interaction strategies for an affective conversational agent. *Presence*, 20(5) :395–411, 2011.
- [7] Kai Sun, Su Zhu, Lu Chen, Siqui Yao, Xueyang Wu, and Kai Yu. Hybrid dialogue state tracking for real world human-to-human dialogues. In *INTERSPEECH*, pages 2060–2064, 2016.
- [8] Ciprian-Octavian Truica, Florin Radulescu, Alexandru Boicea, and Ion Bucur. Performance evaluation for crud operations in asynchronously replicated document oriented database. In *2015 20th International Conference on Control Systems and Computer Science*, pages 191–196. IEEE, 2015.
- [9] Jason D. Williams, Kavosh Asadi, and Geoffrey Zweig. Hybrid code networks : practical and efficient end-to-end dialog control with supervised and reinforcement learning. *CoRR*, abs/1702.03274, 2017.