Les jeux de données en compréhension du langage naturel et parlé : paradigmes d'annotation et représentations sémantiques

Rim Abrougui[1, 2]
(1) Orange Innovation, Lannion, France
(2) Aix-Marseille Université, LIS UMR 7020, Marseille, France
rim.abrougui@orange.com

RÉSUMÉ

La compréhension du langage naturel et parlé (NLU/SLU) couvre le problème d'extraire et d'annoter la structure sémantique, à partir des énoncés des utilisateurs dans le contexte des interactions humain/machine, telles que les systèmes de dialogue. Elle se compose souvent de deux tâches principales : la détection des intentions et la classification des concepts. Dans cet article, différents corpora SLU sont étudiés au niveau formel et sémantique : leurs différents formats d'annotations (à plat et structuré) et leurs ontologies ont été comparés et discutés. Avec leur pouvoir expressif gardant la hiérarchie sémantique entre les intentions et les concepts, les représentations sémantiques structurées sous forme de graphe ont été mises en exergue. En se positionnant vis à vis de la littérature et pour les futures études, une projection sémantique et une modification au niveau de l'ontologie du corpus MultiWOZ ont été proposées.

ABSTRACT

The Challenges of Spoken Language Understanding Datasets : A Study on Annotations and Semantic Representations

Natural and Spoken Language Understanding (NLU/SLU) covers the problem of extracting and annotating the meaning structure from user utterances in the context of human/machine interaction, such as dialogue systems, consisting oftenly of two main tasks : intent detection and slot filling. In this paper, different SLU corpora were studied at a formal and semantic level : their different annotation formats (flat and structured) and ontologies were compared and discussed. With their expressive power maintaining the semantic hierarchy between intents and slots, graph semantic representations were highlighted. In line with the literature and for future studies, a semantic projection and a modification of the ontology of the MultiWOZ corpus were proposed.

MOTS-CLÉS : Compréhension du langage, ontologies, représentation sémantique à plat (BIO), représentation sémantique structurée (graphe).

KEYWORDS: Language Understanding, ontologies, flat semantic representation (BIO), structured semantic representation (graph).

1 Introduction

La compréhension du langage naturel et parlé est un sujet d'étude important dans le cadre des interactions homme-machine. Le domaine comprend plusieurs niveaux d'étude, mais actuellement la tâche de compréhension est principalement axée sur la compréhension de la sémantique globale

des requêtes des utilisateurs et sur l'identification des concepts génériques des mots-clés, à savoir, la détection des intentions et l'identification des concepts (Tur & De Mori, 2011).

Bien qu'il existe plusieurs approches de représentations sémantiques, la majorité des méthodes se base sur la représentation à base de frames sémantiques en utilisant des modèles supervisés pour la classification et l'étiquetage de séquence. Ces modèles qui sont basés sur des réseaux de neurones utilisant des modèles de langage pré-entraînés, ont obtenu des performances élevées sur plusieurs jeux de données SLU qui sont représentés avec un schéma à plat (Béchet & Raymond, 2019) (cf. figure 3).

Cependant, les interactions dans une conversation en conditions réelles sont beaucoup plus complexes. Afin de relever ces défis, il est nécessaire d'avoir d'un côté des corpus d'apprentissage dotés de représentations sémantiques complexes et contextuelles, et de l'autre côté des schémas d'annotation capables de prendre en compte les représentations sémantiques hiérarchiques. De plus, pour construire des modèles de SLU plus robustes et pour les comparer de manière plus précise, il est primordial d'unifier les ensembles de données existants. Cette unification permettra de diversifier les domaines et de fournir plus de connaissances aux systèmes de compréhension du langage qui pourront par conséquent apprendre plusieurs structures sémantiques.

Il existe plusieurs études complètes sur la compréhension du langage naturel et parlé, telles que (Weld *et al.*, 2022) et (Qin *et al.*, 2021), mais cet article explorera plus la question des jeux de données en étudiant leurs ontologies et formats d'annotation et de représentation sémantique. Dans le cadre de la problématique d'unification de toutes les ressources publiques d'apprentissage et d'évaluation des systèmes NLU/SLU, nous avons comparé les différentes ontologies et schémas d'annotation. Nous mettons ainsi en exergue le potentiel des représentations sémantiques structurées qui peuvent être utilisées plus facilement avec le développement des modèles de génération du langage. En s'intéressant particulièrement aux représentations en graphe qui préservent la hiérarchie et le lien sémantique entre les différents labels, nous proposerons une projection sémantique et une modification au niveau de l'ontologie du corpus MultiWOZ2.3.

La présentation des jeux de données et la comparaison de leurs ontologies sont présentées dans la section 2, alors que l'étude des schémas d'annotation et des représentations sémantiques structurées ainsi que nos perspectives sont exposées dans la section 3.

2 Exploration des jeux de données en compréhension du langage

Les jeux de données jouent un rôle crucial dans l'avancement de la recherche dans le domaine de la compréhension du langage. Ils permettent en effet d'entraîner, d'évaluer et de comparer les systèmes SLU. Les corpora disponibles sont variés et peuvent couvrir plusieurs domaines. La façon dont les annotations sont représentées et le choix des labels reflètent une certaine variation au niveau des ontologies ce qui rend difficile l'unification de ces jeux de données. Les schémas de projection et des représentations sémantiques choisis affectant la qualité d'annotation peuvent être aussi différents. Certains corpora sont des conversations complètes multi-domaines et/ou multi-intentions, tandis que d'autres ne contiennent que des simples requêtes. Malgré cette diversité, un corpus large avec des conversations complètes entièrement annotées en logique SLU selon une ontologie générique applicable à toutes les données est difficile à trouver, ce qui rend les travaux sur l'exploitation de l'histoire conversationnelle et du contexte plus difficile.

Nous pouvons trouver également plusieurs types de collectes de ces corpus. L'une de ces méthodes est appelée la méthode «Wizard-Of-OZ» où les participants interagissent en temps réel avec un système qu'ils croient être autonome, mais il est contrôlé en réalité par un opérateur humain invisible. Cette méthode est plus fréquente puisqu'elle permet de collecter des données de manière contrôlée, avec une grande variété de scénario de dialogues. Les transcriptions de la parole à l'aide de la reconnaissance automatique de la parole (ASR), sont un autre moyen de collecter des données. La qualité des transcriptions va dépendre des systèmes ASR mais cette méthode permet de collecter d'une façon plus rapide les énoncés dans des conditions réelles et qui vont être vérifiés et annotés manuellement ou d'une manière semi-automatique. Nous trouvons enfin des méthodes plus automatisées qui consistent à synthétiser les conversations à l'aide de modèles de langage ou de modèles de génération de texte. Cette méthode est la plus rapide mais elle est artificielle car elle ne peut pas refléter toutes les variations linguistiques dans la parole humaine et peut avoir des problèmes d'hallucination. Cette section présentera certains jeux de données publiés pour la tâche SLU, ainsi que leurs ontologies et leurs méthodes d'annotation.

2.1 Les ensembles de données pour les tâches de compréhension du langage naturel et parlé

Jeux de données	Langues	Modes	#énoncés/dialogues	#labels
ATIS (**Hemphill *et al.*, 1990**)	en	requêtes	4978	17 intentions 84 slots
Frames (**Asri *et al.*, 2017**)	en	dialogues	1369	20 actes 16 slots
Massive (**FitzGerald *et al.*, 2022**)	multilingues 51 langues	requêtes	19521 par langue	60 intentions 55 slots
MEDIA (**Devillers *et al.*, 2004**)	fr	dialogues	1250	83 slots 19 spécifieurs
mTOD (**Schuster *et al.*, 2019**)	multilingues 3 langues	requêtes	43000	12 intentions 11 slots
mTOP (**Li *et al.*, 2020**)	multilingues 6 langues	requêtes	10000	117 intentions 78 slots
MultiDoGo (**Peskov *et al.*, 2019**)	en	dialogues	15000 annotés	85 intentions 73 slots
MultiWOZ (**Budzianowski *et al.*, 2018**)	en	dialogues	10438	32 actes 27 slots
M2M (**Shah *et al.*, 2018**)	en	dialogue	3000	15 actes 12 slots
SNIPS (**Coucke *et al.*, 2018**)	en	requêtes	14484	7 intentions 39 slots
TOP (**Gupta *et al.*, 2018**)	en	requêtes	44783	25 intentions 36 slots
The restaurant-8K dataset (**Coope *et al.*, 2020**)	en	requêtes	8198	5 slots 5
VocaDOM (**Portet *et al.*, 2019**)	fr	requêtes	4610	7 intentions 12 slots

TABLE 1 – Tableau récapitulatif des ensembles de données en NLU/SLU

Il existe de nombreux corpora SLU disponibles publiquement, chacun ayant ses propres caractéristiques et domaines d'application. Les tableaux 1 et 2 synthétisent les caractéristiques de ces données et nous détaillons ci-dessous chaque corpus.

Jeux de données	Multi-Domaines	Multi-Intents	Inter-Domaines	Annot. contextuelles	Annot. à plat	Annot. structurée ou semi-structurée
ATIS (Hemphill *et al.*, 1990)					X	
Frames (Asri *et al.*, 2017)				X		X
Massive (FitzGerald *et al.*, 2022)	X				X	
MEDIA (Devillers *et al.*, 2004)					X	
mTOD (Schuster *et al.*, 2019)	X				X	
mTOP (Li *et al.*, 2020)	X				X	X
MultiDoGo (Peskov *et al.*, 2019)	X	X				X
MultiWOZ (Budzianowski *et al.*, 2018)	X	X	X	X		X
M2M (Shah *et al.*, 2018)						X
SNIPS (Coucke *et al.*, 2018)	X				X	
TOP (Gupta *et al.*, 2018)	X					X
The restaurant-8K dataset (Coope *et al.*, 2020)				X	X	
VocaDOM (Portet *et al.*, 2019)					X	

TABLE 2 – Tableau récapitulatif des caractéristiques des jeux de données en NLU/SLU

2.1.1 Jeux de données avec des requêtes simples

1. Ressources mono-lingues :
   (a) **ATIS** (Hemphill *et al.*, 1990) : Le corpus ATIS (Air Travel Information System) est l'un des corpus SLU les plus utilisés. Il contient des informations sur des compagnies aériennes et des commandes pour réserver des vols. La première version a été collectée suivant l'approche «`Wizard-Of-OZ`», mais les auteurs ont utilisé des transcriptions ASR pour les autres versions (Dahl *et al.*, 1994). Les premières annotations de ce corpus ont été effectuées à l'aide d'une requête SQL, puis ont été transférées au niveau global sous forme d'intentions, ainsi qu'au niveau mot sous forme des concepts (Béchet & Raymond, 2018). Ce corpus est en anglais et contient 4978 énoncés annotés en frame avec 17 intentions et 84 concepts.

   (b) **Frames** (Asri *et al.*, 2017) : Il s'agit d'un corpus avec des interactions humain-humain en anglais et qui contient des informations sur les réservations d'hôtel. Il a été publié pour encourager les recherches sur les systèmes conversationnels textuels. La notion de «mémoire» et l'exploitation de l'histoire conversationnelle ont été les premières questions abordées, où les auteurs ont rajouté à la tâche NLU, la tâche de "suivi des frames sémantiques". Des références et des identifiants des frames sémantiques ont été donc rajoutés aux annotations en acte de dialogue et en slot-valeur. Ce corpus peut être utilisé dans les tâches de compréhension et dans les tâches de suivi d'état de dialogue. 1369 dialogues ont été collectés suivant l'approche «`Wizard-Of-OZ`» . Les énoncés au niveau utilisateurs et systèmes ont été annotés avec 20 types d'acte de dialogue et 16 types de slot.

   (c) **MEDIA** (Devillers *et al.*, 2004) : Le corpus MEDIA contient des informations touristiques en français. Il a été collecté suivant l'approche «`Wizard-Of-OZ`». 1250 dialogues

ont été transcrits et annotés manuellement suivant une ontologie très riche au niveau des énoncés de l'utilisateur. Nous retrouvons 83 concepts de base regroupés dans un dictionnaire sémantique et 19 "spécifieurs" pouvant leur être associés. En lien avec le corpus MEDIA, PORT-MEDIA (Lefevre *et al.*, 2012) a été publié en français et en italien. Les méthodes de collecte étaient similaires en rajoutant des scénarios de dialogue variés. Les annotations étaient semi-automatiques à partir des systèmes de compréhension entraînés sur MEDIA suivies d'une vérification et correction manuelle. La version italienne a été générée par des traductions automatiques de la version française. Les questions de la complexité sémantique et les différents niveaux hiérarchiques ont été creusées et traitées sur une base linguistique solide.

(d) **SNIPS** (Coucke *et al.*, 2018) : SNIPS est un corpus en anglais qui contient plusieurs domaines différents. Les données ont été collectées par des transcriptions ASR suivi d'une annotation et vérification manuelles. Il contient 7 intentions et 39 concepts. La même version de ce corpus a été publiée en d'autres langues comme le français et l'allemand. SNIPS a été l'origine d'un autre corpus, **Almawave SLU** (Bellomaria *et al.*, 2019), le premier ensemble de données en italien pour les expériences SLU. Il a été généré d'une manière semi-automatique en traduisant les énoncés et les labels et en remplaçant les entités ouverts (comme les noms de restaurants et des livres) par des références italiennes. La vérification et la correction manuelles a été effectuée pour les énoncés.

(e) **TOP** (Gupta *et al.*, 2018) : Ce corpus a été publié pour étudier les problématiques sémantiques plus complexes. Les auteurs ont introduit une représentation hiérarchique appelée "*Task Oriented Parsing*" (TOP) pour les systèmes de dialogue basés sur des intentions et des concepts. Les énoncés ont été collectés par *crowd-sourcing* et annotés par deux annotateurs, un troisième annotateur peut intervenir en cas de désaccord. 44783 annotations avec 25 intentions et 36 slots ont été obtenues. Une version étendue du corpus avec 6 domaines supplémentaires a été publié dans **TOPv2** (Chen *et al.*, 2020).

The restaurant-8K dataset (Coope *et al.*, 2020) : Pour renforcer le travail d'extraction de concepts dans le cadre de dialogues, ce jeu de donnée qui comprend des conversations d'un système de réservation de restaurant, a été introduit. 8198 énoncés d'utilisateurs réels interagissant avec un système de dialogue déployé dans le domaine de la réservation de restaurants ont été annotés d'une manière contextuelle indiquant quels concepts ont été demandés par le système. Les réponses de système ne sont pas incluses dans l'ensemble des données, et il n'y a que 5 concepts.

VocaDOM (Portet *et al.*, 2019) : Afin de soutenir les tâches dans le cadre des systèmes *"Smart Home*" comme l'identification du locutaire, la reconnaissance de la parole et les tâches SLU, ce corpus a été publié en rassemblant des interactions dans des conditions réelles de 11 participants dans une maison intelligente, la méthode «`Wizard-Of-OZ`» était la base de ce protocole. Les enregistrements ont été transcrits et annotés manuellement par des intentions et des concepts. Au total, le corpus contient 4610 énoncés en français étiquetés par 7 intentions et 12 concepts. Dans (Desot *et al.*, 2018), un jeu de donnée synthétiques du même domaine a été généré automatiquement à partir du corpus VocaDOM.

2. Ressources multi-lingues :

(a) **Massive** (FitzGerald *et al.*, 2022) : Massive est un corpus multilingue qui contient des requêtes appartenant à 18 domaines. Sa publication a été motivée par le manque des jeux de données en plusieurs langues pour évaluer les modèles multilingues. Le corpus

SLURP (Bastianelli *et al.*, 2020), publié pour développer un assistant robotique personnel à domicile et pour des expériences SLU *End-to-end*, est la version d'origine du Massive. La version du corpus SLURP disponible publiquement est textuelle en anglais, elle a été collectée par les travailleurs de «Mechanical Turk» (AMT) et annotée manuellement au niveau "scénario" (domaine), "action" (Intention) et "entités" (concepts). Des traducteurs professionnels ont traduit les énoncés du corpus en 51 langues et ont également vérifié les frontières des concepts sur les tokens, aboutissant à la création du corpus Massive. Ce dernier est considéré comme une grande source des intentions (60) et de concepts (55) vu la diversité des domaines.

(b) **mTOD** (Schuster *et al.*, 2019) : Il s'agit d'un ensemble de données multilingue qui permet d'étudier les méthodes d'apprentissage par transfert inter-linguistique. Ce corpus offre l'opportunité d'étudier les modèles sémantiques inter-langues et constitue le premier ensemble de données parallèles pour une tâche d'étiquetage de mots qui a été annoté selon les mêmes guides d'annotation dans plusieurs langues. Les auteurs ont collecté 43000 énoncés en anglais dans les domaines *ALARM*, *REMINDER*, et *WEATHER*, et ils ont demandé à des anglophones natifs de proposer des labels d'intentions utilisées par deux annotateurs pour étiqueter les énoncés et les valeurs par des concepts. Cette annotation a été vérifiée ensuite par un troisième annotateur. Des locuteurs natifs en espagnol et en thaï ont traduits les énoncés qui ont été aussi annotés par deux annotateurs. Il contient au total 12 types d'intentions et 11 concepts

(c) **mTOP** (Li *et al.*, 2020) : En creusant la même problématique de la sémantique compositionnelle mise en évidence dans le corpus TOP (Gupta *et al.*, 2018) et en suivant la même logique de représentation hiérarchique, le corpus mTOP, a été publié. Cet ensemble de données est le premier qui contient des représentations sémantiques compositionnelles qui permettent l'annotation des requêtes imbriquées. Il a été publié avec les deux versions d'annotation : une plate et une autre compositionnelle. Les auteurs ont commencé par collecter une version en anglais des données, suivant la même approche dans (Gupta *et al.*, 2018), qui est traduite ensuite par des traducteurs professionnels. Le corpus mTOP est plus grand que TOP où nous avons 100.000 exemples avec 6 langues différentes, 11 domaines, 117 intentions et 78 concepts. En outre, une version parlée (**STOP**) a été publiée dans (Tomasello *et al.*, 2023) à partir de **TOPv2** pour encourager les recherches sur les approches end-to-end tout en focalisant sur les problématiques des requêtes compositionnelles.

2.1.2 Jeux de données conversationnels

1. Ressources mono-lingues :

(a) **MultiDoGo** (Peskov *et al.*, 2019) : Cet ensemble de données a été collecté par «crowd-sourcing» dans le cadre du progrès des assistants virtuels et du manque des données pour leur développement. Ce corpus est composé par 81000 conversations, dont 15000 ont été annotées avec 6 domaines différents, 85 intentions et 73 slots. Dans le corpus disponible publiquement, les énoncés systèmes ne sont pas annotés. L'article présentant ces données a mis en valeur la possibilité d'avoir des multi-intentions en montrant que les annotations ont été réalisées par des experts selon deux niveaux : au niveau des tours des dialogues et au niveau des phrases, afin de garder l'ordre entre les énoncés coordonnées et leurs intentions. Toutefois, dans la version publiée, nous ne trouvons pas souvent cette illustration et nous pouvons même perdre le lien entre les différentes

intentions et leurs concepts.

(b) **MultiWOZ** (Budzianowski *et al.*, 2018) : Il s'agit d'un corpus de dialogue multi-domaines en anglais, à grande échelle, souvent utilisé pour plusieurs tâches, notamment le suivi de l'état du dialogue, la politique de dialogue et les tâches de génération de dialogue. Il a été collecté à partir de la méthode «`Wizard-Of-OZ`» via un «`crowd-sourcing`». La première version de ce corpus a été publiée dans le but de faciliter la construction de systèmes de dialogue supervisés. Chaque énoncé dans les dialogues est annoté avec une séquence d'acte de dialogue. Cependant, la première version comporte des erreurs d'annotations, surtout au niveau de l'utilisateur, puisqu'elles ont été effectuées automatiquement à partir des annotations système (Eric *et al.*, 2019). Plusieurs versions ont été produites pour corriger ces erreurs et simplifier le format des annotations. Les versions MultiWOZ2.2 (Zang *et al.*, 2020) et MultiWOZ2.4 (Ye *et al.*, 2021) sont mieux adaptées aux tâches de suivi de l'état du dialogue, tandis que la version MultiWOZ2.3 (Han *et al.*, 2020) a des annotations utilisateur plus précises. Le corpus contient 7 domaines, 32 actes de dialogue et 27 slots.

(c) **M2M** (Shah *et al.*, 2018) : M2M est une fusion de 2 données contenant des dialogues en anglais pour la réservation des restaurants et des tickets de cinémas. Les méthodes de collecte et de l'annotation ont été réalisées d'une manière automatique où un développeur de dialogue fournit un scénario et les chatbots génèrent des tours de conversations en les annotant par des actes de dialogue et par des slots. Ce processus était répété jusqu'à la fin des tours de dialogue soit par un acte "bye" soit en atteignant un seuil maximum de tours. Au total, nous avons 3000 dialogues annotés avec 15 actes de dialogues et 12 slots.

Dans la partie suivante, nous allons nous focaliser sur l'analyse de certains ensembles de données au niveau de leurs ontologies et schémas sémantiques.

2.2 Ontologies et Annotations

L'annotation sémantique repose généralement sur des ontologies basées sur des connaissances linguistiques permettant de définir des liens hiérarchiques entre les entités (Ma *et al.*, 2009). Dans les tâches de compréhension du langage, les ontologies permettent de décrire le lien sémantique entre les domaines, les intentions ou les actes de dialogue et leurs concepts (Loos, 2006). Ces ontologies se varient selon les schémas et les règles d'annotation des corpus, mais parfois, nous pouvons trouver la même logique surtout lorsque le schéma d'annotation est limité à un cadre sémantique simple. En d'autres termes, dans des jeux de données comme le cas d'ATIS (Hemphill *et al.*, 1990), de SNIPS (Coucke *et al.*, 2018), de Massive (FitzGerald *et al.*, 2022) et de mTOD (Schuster *et al.*, 2019), chaque énoncé est labélisé par une seule intention, la notion de multi-intentions ou le croisement entre les domaines (inter-domaines) sont donc absents pour ces corpora.

En plus, les données citées peuvent avoir des domaines en commun et donc des similarités au niveau des concepts. Par exemple, le corpus ATIS n'a qu'un seul domaine (réservation de vol) qui peut se croiser avec le domaine "airline" dans le corpus MultiDoGo (Peskov *et al.*, 2019). SNIPS a aussi plusieurs domaines en commun avec TOP (Gupta *et al.*, 2018) («`Restaurant, Weather, Music`»...). Cependant, les différentes façons d'exprimer les intentions et le choix des concepts entraînent une grande diversité au niveau des ontologies, ce qui rend leur unification assez problématique. Les intentions dans ATIS sont des noms simples, par contre dans mTOD et SNIPS elles sont composées

d'un acte de dialogue avec le domaine («Show_alarms», «BookRestaurant»). Les actes de dialogue dans MultiWOZ sont aussi composés par le domaine associé à l'acte («Hotel-Inform»), mais le choix des actes est associé à la sémantique derrière les concepts plus qu'au sens global des énoncés. Autrement dit, les concepts comme «Phone» et «Car» pour le domaine «Taxi» sont toujours associés à l'acte «Request».

Les concepts peuvent être aussi variés au niveau de leur composition, ils peuvent se représenter comme une seule entité («country») ou une entité composée («party_size_description»). Nous avons remarqué aussi que dans toutes ces données les concepts sont plus compositionnels et reliés à leurs domaines. Dans le corpus Massive, on a par exemple les slots «sport_type», «drink_type» et «alarm_type», à l'antipode de MultiWOZ qui n'a que le concept «Type» partagé par les domaines «Attraction» et «Hotel». Dans le corpus ATIS, les concepts présentent un niveau plus haut au niveau de sa composition où les prépositions peuvent faire partie des slots (comme «from» dans «fromloc.city_name»), ou nous pouvons même trouver des concepts imbriqués (le slot «depart_time.period_of_day» est composé par le slot «time» et le slot «period_of_day»).

Le corpus MEDIA (Devillers *et al.*, 2004) est par ailleurs assez particulier au niveau de son schéma d'annotation. L'ontologie a été basée sur un niveau sémantique haut qui essaie de relever le lien hiérarchique des labels sémantiques. Les frames sémantiques ne se composent pas que par des paires de slot-valeur, mais aussi par un spécifieur qui définit les relations entre les entités. Une annotation des modes a été aussi effectuée et attachée aux concepts. Ainsi, la représentation hiérarchique a été recomposée par la combinaison des "spécifieurs" et des concepts.

Il convient également de noter que ces ensembles de données varient considérablement en termes de complexité linguistique. Dans l'article (Bechet *et al.*, 2022), divers phénomènes linguistiques qui peuvent impacter les performances des systèmes SLU sont observés. Certains corpora ne reflètent pas les caractéristiques des interactions dans des conditions réelles, tandis que d'autres sont plus difficiles pour les modèles. L'approche proposée pour évaluer la qualité et comparer les corpora, comme décrite dans (Bechet *et al.*, 2022) et (Béchet & Raymond, 2019), peut contribuer à la question d'unification des données.

En ce qui concerne la manière des projections des frames sémantiques, le format d'annotation "BIO" à plat (cf. tableau 3) est souvent le paradigme le plus utilisé, notamment pour les données avec de simples requêtes, afin de faire un étiquetage de séquence facilement avec les modèles pré-entraînés à base de Transformers de type Bert (Devlin *et al.*, 2018). Néanmoins, ce paradigme est limité si nous voulons passer à la compréhension des énoncés plus complexes en exploitant le contexte de la conversation. Nous avons remarqué les limites de ce paradigme avec le corpus MultiWOZ où un seul énoncé peut avoir plusieurs domaines ou plusieurs intentions (cf. 4). Les annotations d'origine de ce corpus se représentent sous un format plus structuré en format `json` où nous pouvons trouver le lien entre les différents concepts et leurs intentions. Une suggestion d'une annotation à plat a été proposée dans (Lee *et al.*, 2019), mais les difficultés des annotations en contexte, notamment pour les concepts sans informations d'empans, n'ont pas été entièrement résolues. Ces entités sont justement définies comme des "slots de catégories", où leurs valeurs se trouvent dans les énoncés précédents, ou bien implicitement dans le sens global de l'énoncé. Les slots `"Parking"` et `"Post"` dans la table 4 sont un exemple des concepts de catégories.

Les différents paradigmes d'annotation et les différentes hiérarchies sémantiques des ontologies mettent en valeur la difficulté de représenter les données d'une manière structurée où le contexte peut-être bien exploité. Dans la section suivante, cette question sera creusée où nous allons nous

Enoncés	[CLS]	I'm	traveling	to	dallas	from	philadelphia
Annotation	Flight	O	O	O	B-toloc.city_name	O	B-fromloc.city_name

TABLE 3 – Annotation à plat en BIO du corpus ATIS :"B" pour «Begining», "I" pour «Inside» et "O" pour «Outside»

Enoncés	**Annotation en acte de dialogue**
◇ I won't have a car, so parking isn't important	`"Hotel-Inform": [["Parking", "no"]]`
◇ Can I have the postcode for the attraction, I also need a Taxi	`"Attraction-Inform": [["Post", "?"]], "Taxi-Inform": [["none", "none"]]`

TABLE 4 – Exemples d'énoncés annotés dans le corpus MultiWOZ2.3

Enoncés	take grandma Jane off the call
Annotation	`[IN:update_call [SL:contact_removed [IN:get_contact [SL:type_relation grandma ] [SL:contact Jane ]]]]`

TABLE 5 – Exemple d'annotation dans le corpus mTOP

intéresser aux différentes motivations qui sous-tendent l'utilisation de représentations structurées des étiquettes sémantiques des données pouvant être une solution de certaines limites des annotations à plat.

3 Projections des annotations et représentations sémantiques structurées dans le NLU/SLU

Nous avons présenté dans la section précédente les différences et les similitudes entre les jeux de données utilisés pour les tâches de compréhension du langage au niveau de leurs ontologies et leurs schémas d'annotation. Dans cette section, nous allons retracer d'une manière générale l'historique des différentes représentations sémantiques utilisées en compréhension du langage, ainsi que les enjeux liés aux représentations actuelles dans le contexte des systèmes de dialogue. Nous étudierons ensuite les projections structurées en illustrant les problématiques liées aux différents schémas à plat, ainsi que le potentiel des formats structurés, en particulier ceux basés sur la méthode des graphes.

3.1 Représentations sémantiques

Les interprétations sémantiques peuvent être considérées comme un processus de traduction, réalisé par un parseur sémantique, entre les mots d'une phrase et les représentations sémantiques du langage, comme montré dans (Dinarelli, 2010). Les représentations sémantiques permettent la modélisation des énoncés et de leurs interprétations sémantiques pour les machines. Elles peuvent être sous forme de logique formelle (Zettlemoyer & Collins, 2012), des frames sémantiques (Dinarelli *et al.*, 2009) ou encore des graphes sémantiques (Banarescu *et al.*, 2013).

Avec le progrès des systèmes de dialogue et des modèles modernes basés sur des approches de

statistiques et de probabilités, les interprétations sémantiques en SLU sont principalement basées le plus souvent sur l'identification des intentions ou des actes de dialogue et des slots. En outre, ces annotations notamment au niveau des slots sont très similaires aux annotations par frame sémantique, comme dans FrameNet (Baker *et al.*, 1998), dans la mesure où la représentation SLU est basée sur des attributs, qui sont des unités sémantiques, instanciées par séquences de mots. Contrairement aux frames, les attributs en SLU n'ont pas besoin d'expliciter les relations sémantiques entre les éléments de la phrase (Dinarelli, 2010). Au niveau concepts, les étiquettes consistent à identifier les éléments clés de l'énoncé, comme les entités nommées, les dates ou les destinations. Quant aux intentions, les annotations sont utilisées pour aider les systèmes SLU à comprendre le but global de l'utilisateur. Nous pouvons trouver ainsi des annotations en acte de dialogue pour mieux représenter les intentions. Par ailleurs, des recherches ont été menées pour représenter les attributs sémantiques du dialogue sous forme d'une ontologie abstraite, générique et structurée en exploitant les représentations AMR pour l'analyse sémantique du dialogue. (Bonial *et al.*, 2020).

Il est important ainsi de réfléchir à la question de la projection des informations sémantiques pour les traduire en entrées exploitables par les modèles. L'approche la plus courante pour la tâche de compréhension du langage repose sur les approches de l'étiquetage de séquence. La projection à plat en format BIO facilite cette tâche notamment pour les approches jointes pour prédire les intentions et les concepts simultanément. Comme il est montré dans l'exemple 3, ces approches associent généralement l'intention au token de classification globale [CLS] et détectent les concepts sur chaque token concerné avec une étiquette B ou I lorsque cela est possible.

Bien que la projection basée sur le format BIO soit utile pour l'utilisation des modèles à base de Transformers, elle ne permet pas de tirer parti des structures sémantiques hiérarchiques et imbriquées. Ainsi, des recherches sur des annotations plus structurées ont été étudiées, notamment avec l'utilisation des approches de séquence à séquence (*seq2seq*). Les chercheurs dans (Li *et al.*, 2020) ont proposé une représentation composée découplée (cf. 5) pour représenter des intentions imbriquées dans les slots. De même les expériences dans (Hu *et al.*, 2022) présentent le NLU comme une tâche de génération des graphes composés par des nœuds pour les labels. Dans la section suivante, nous allons montrer les différentes motivations et illustrations des schémas de représentation structurée.

3.2 Les représentations structurées des frames sémantiques

Selon (Devillers *et al.*, 2004), la représentation sémantique des annotations d'un corpus a été définie comme un moyen pour représenter les frames sémantiques d'une manière générique et complète selon la tâche mais qui permet aussi l'annotation des corpora larges d'une manière simple. Par conséquent, les schémas de représentation à plat ont été le centre des annotations dans la majorité des copora publics. Cependant, les représentations hiérarchiques sont plus expressives et permettent le lien entre les sous-structures (Tur & De Mori, 2011).

Dans la section 2.2 nous avons présenté quelques tentatives de projection des annotations d'origine du corpus multi-domaines MultiWOZ au paradigme à plat dans (Lee *et al.*, 2019). Comme nous remarquons dans l'exemple 6, le schéma de représentation a repris l'idée de compositionnalité des labels notamment pour les concepts de catégorie, où l'ensemble de l'acte de dialogue, le concept et sa valeur ont été projetés au niveau global `[CLS]`. Il est important de souligner cependant que cette projection demeure limitée si l'on veut prédire des représentations plus complexes.

En outre dans (Gupta *et al.*, 2018) le paradigme à plat a été remis en question pour les requêtes

Enoncés	[CLS]	I	need	parking
Annotation	Hotel-Inform+Parking*yes	O	O	O

TABLE 6 – Annotation à plat en BIO du corpus MultiWOZ

plus complexes. En effet le corpus TOP est modélisé d'une manière compositionnelle qui autorise les intentions imbriquées, il s'agit en effet d'une représentation hiérarchique similaire aux arbres syntaxiques. Vu la complexité de la représentation et pour faciliter l'utilisation du même schéma en plusieurs langues, (Aghajanyan *et al.*, 2020) ont proposé une extension de la représentation compositionnelle en représentation découplée qui a été utilisée dans (Li *et al.*, 2020). La figure 1 illustre la projection : le premier niveau de l'arbre correspond à l'intention, qui peut inclure un ou plusieurs concepts. Ces derniers peuvent à leur tour comporter des intentions ou une séquence de mot comme une valeur. En somme, les auteurs ont démontré que cette projection est un compromis entre le paradigme traditionnel à plat et la représentation en logique formelle. De même, les expériences faites dans (Cheng *et al.*, 2020) s'inscrivent dans le même cadre de la sémantique compositionnelle en affirmant que la compositionnalité peut simplifier la compréhension pour faciliter la tâche de suivi d'état de dialogue.

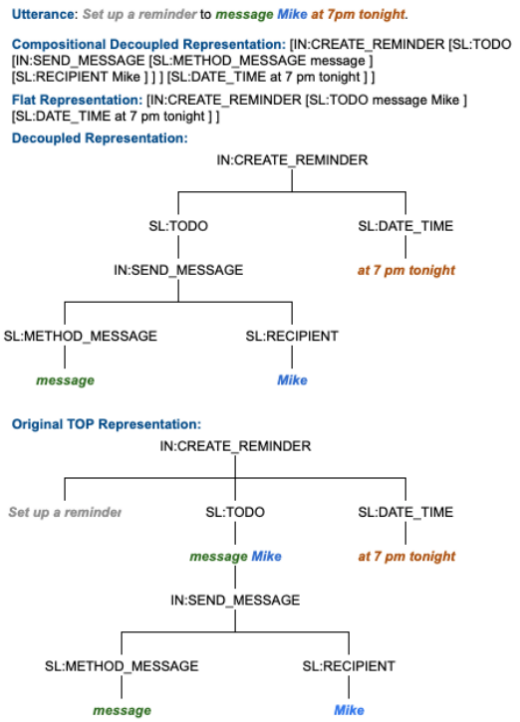

FIGURE 1 – Représentation compositionnelle découplée vs représentation plate dans mTOP (Li *et al.*, 2020)

De surcroît, les limitations du paradigme à plat et les motivations d'une représentation plus structurée ont été discutées dans (Hu *et al.*, 2022). L'article a proposé «DMR» une représentation en graphe qui comprend des nœuds d'Intention, de slots, des opérateurs indiquant les coréférences et la conjonction,

et des mots-clés pour quelques éléments spéciaux en sémantique comme la négation. Une définition d'une nouvelle ontologie du domaine «`Fast Food`» du corpus MultiDoGo et une ré-annotation structurée qui lie les tours de dialogue permettant les annotations en contexte ont été l'objet de cet article. Des notions comme la "quantification" et "les adjectifs modificateurs" ont été ainsi soulignées. Un exemple de leur représentation est dans la figure 2.

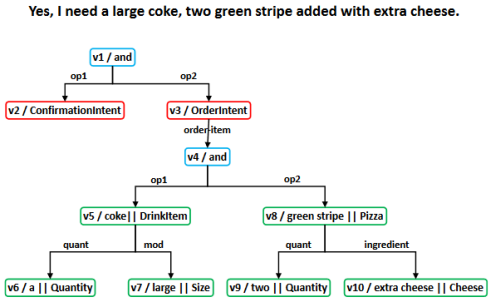

FIGURE 2 – Représentation en graphe dans DMR (Hu *et al.*, 2022)

Une proposition similaire à (Hu *et al.*, 2022) a été suggérée dans (Abrougui *et al.*, 2022). Cette proposition illustre des projections effectuées sur les annotations d'origine du corpus MultiWOZ2.3 sans qu'il soit nécessaire de modifier l'ontologie. Tel qu'indiqué dans la figure 3 Les actes de dialogue et les slots-valeurs sont transposés comme des nœuds, tandis que les slots sont représentés sous forme d'arcs reflétant la hiérarchie entre les labels. Les cas complexes, tels que les multi-intentions ou les intentions imbriquées peuvent être projetés dans ce format grâce à l'encodage Penman (Kasper, 1989), également utilisé dans les analyses AMR (Banarescu *et al.*, 2013).

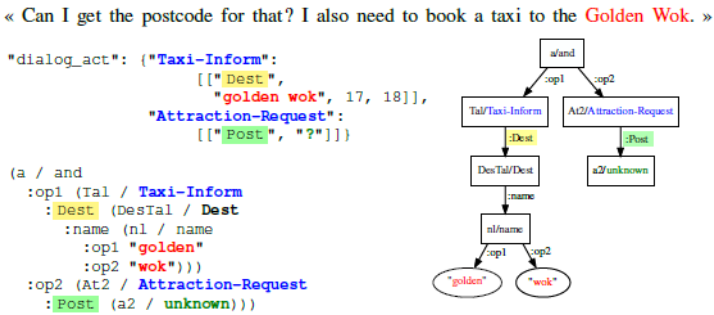

FIGURE 3 – Représentation en graphe et encodage Penman dans (Abrougui *et al.*, 2022)

Avec le développement des modèles de réseaux de neurones et l'essor des approches *seq2seq*, les représentations structurées pour les annotations sont devenues facilement exploitables par les systèmes NLU. Dans le même contexte, nous présentons nos perspectives et projets de recherche dans la partie suivante.

3.3 NLU avec une annotation structurée sur le corpus MultiWOZ

Les schémas d'annotation structurés et hiérarchiques offrent la possibilité d'étudier d'une façon plus approfondie une sémantique complexe et compositionnelle. Nous avons choisi donc de travailler sur le corpus MultiWOZ présentant des défis intéressants en raison de sa complexité (multi-domaines, croisement entre les différentes intentions et concepts). Nous avons choisi de travailler en particulier sur la version 2.3, car elle offre des annotations utilisateur plus précises que les autres versions. Cependant, nous avons constaté que les annotations de cette version comportent des erreurs selon les règles de la logique NLU. Pour remédier à cela, notre objectif est de corriger et d'enrichir l'ontologie du corpus et de proposer un modèle de représentation qui permet une annotation contextuelle plus précise liant les tours de dialogue entre eux.

3.3.1 Corrections des annotations

Nous avons constaté dans la section 2 que les premières annotations au niveau utilisateurs de MultiWOZ ont été générée automatiquement. Bien que les auteurs dans MultiWOZ2.3 aient apporté des corrections, des précisions manquent encore. Nous avons donc commencé à examiner les conversations en identifiant une liste des mots-clés, tels que «`I want to travel from`» ou «`I want to arrive by`», qui permettent de sélectionner un ensemble d'énoncés à vérifier et à corriger. Notre objectif dans cette étape est de garantir la cohérence entre les actes de dialogue, les concepts et les valeurs sans modifier l'ontologie de base.

En d'autres termes, l'exemple dans le tableau 7 doit être corrigé comme `"Hotel-Request": [["Internet", "free"]]` puisque l'utilisateur demande une information spécifique sur l'accès gratuit à Internet. Toutefois, la combinaison de la valeur `"free"` avec l'acte `"Request"` n'existe pas dans l'ontologie du corpus. Les actes dans MultiWOZ sont choisis en fonction du type des concepts, (slots-valeur en cas des concepts de catégorie), plutôt que du sens global de l'énoncé. Ainsi dans cette ontologie, la valeur de catégorie `"free"` est associée au concept `"Price"`, qui est lui-même associé à l'acte `"Inform"`. Cependant, le slot `"Price"` n'est utilisé que pour définir les prix des hôtels et des restaurants. Dans le cas du slot `"Internet"`, on a 4 valeurs principales : `"yes"` `"no"` `"dontcare"` associées à l'acte `"Inform"` et la valeur `"?"` associée à l'acte `"Request"`. Etant donné que l'énoncé dans l'exemple 7 est une simple demande d'information, nous nous limitons donc à corriger l'annotation en `"Hotel-Request": [["Internet", "?"]]`.

Stratégie	Recherche de l'expression `"do they offer free wifi ?"`
Type de correction	semi-automatique
Annotation d'origine	`"Hotel-Inform": [["Internet", "yes"]]`
Correction	`"Hotel-Request": [["Internet", "?"]]`

TABLE 7 – Exemple de correction dans MultiWOZ2.3

3.3.2 Projection structurée

Dans (Abrougui *et al.*, 2022) nous trouvons une projection structurée des intentions, concepts et valeurs du MultiWOZ sans effectuer aucune modification (cf. figure 3). L'avantage de cette

représentation est sa capacité de projeter les jeux de données standards comme ATIS, et les requêtes les plus complexes comme dans TOP et dans MultiWOZ. Dans cette étape nous avons deux étapes. Tout d'abord, nous visons à rajouter de nouveaux labels à l'ontologie du corpus comme l'acte `"Confirmation"`, ou comme les adverbes représentés par un concept de catégorie, comme il est montré dans l'exemple 8. Nous réfléchissons aussi à la question si l'acte doit être associé ou dissocié du domaine.

Enoncé	Can you also give me some information about Finches Bed and Breakfast ? We 're thinking of staying there .
Annotation d'origine	`"Hotel-Inform": [["Name", "finches bed and breakfast"]]`
Proposition d'annotation	`"Hotel": [["Inform-Name", "finches bed and breakfast"],["Request-Info", "?"], ["Modifier", "maybe"]]`

TABLE 8 – Proposition d'une nouvelle annotation du corpus MultiWOZ2.3

Notre objectif en second temps est de reprendre cette structure en rajoutant des annotations de coréférence en liant les antécédents et les anaphores avec les variables utilisées dans la notation Penman. La figure 4 ci-dessous illustre ce projet. En examinant l'énoncé, nous remarquons la présence de deux actes de dialogues distincts qui partagent le slot `"Area"` clairement indiqué par l'utilisation de l'adverbe `"there"`. La représentation sémantique de cette structure avec un format à plat serait particulièrement difficile. En outre, il convient de souligner la problématique de l'implicite soulevée par l'expression `"I have a car"`, qui fait référence au concept de catégorie `"Parking"` et à sa valeur normalisée `"yes"`. Afin de faciliter la représentation de ces entités et de leurs liens, l'utilisation du format penman basé sur les variables (comme `"a1"` dans la figure) s'avère être un choix judicieux.

Notre objectif est en effet de faire une tâche NLU mais en couvrant toutes les informations possibles, comme l'annotation de la coréférence.

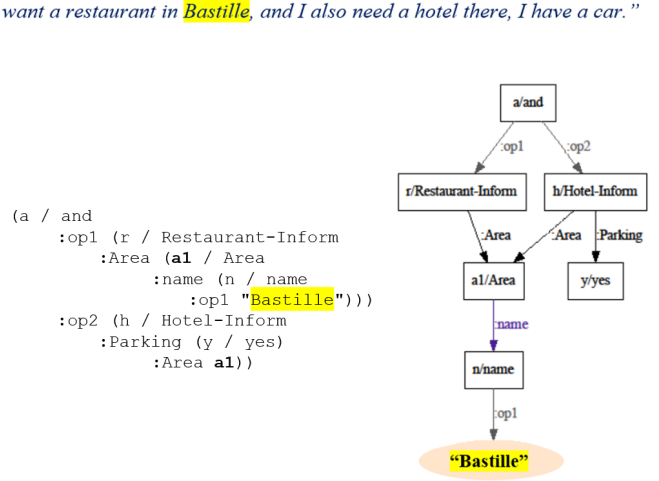

FIGURE 4 – schéma structuré pour la coréférence

3.3.3 Pistes d'amélioration

Il est vrai que les schémas de représentation structurés sont considérés comme des outils efficaces pour représenter les données avec des informations sémantiques complexes, mais la sémantique sous-jacente des étiquettes est tout aussi importante. Les auteurs dans (Athiwaratkun *et al.*, 2020) ont souligné cette question en expliquant que les modèles de langage génératifs offrent un moyen naturel d'incorporer le sens des étiquettes dans les tâches de compréhension. Ils ont représenté la sortie en une séquence augmentée qui contient la séquence d'entrée avec leurs labels. Ils ont nettoyé les labels de tous les symboles et les ont représentés sous leur forme de langage naturel (par exemple, "AddToPlaylist" est devenu "Add To Playlist") afin de mieux exploiter les capacités des modèles de langage génératifs à comprendre le langage naturel.

Par ailleurs, l'unification des ensembles de données SLU reste un défi dans ce domaine, car les ontologies et les noms des concepts varient considérablement. Néanmoins, si des noms de concepts sont communs et que les modèles génératifs sont capables de comprendre le langage naturel, il est possible d'unifier ces jeux de données plus facilement sans effectuer de nombreuses modifications au niveau de leurs ontologies. Nous avons testé cette hypothèse en examinant les données SNIPS et MultiWOZ.

En effet, SNIPS a un domaine en commun avec MultiWOZ, qui est "Restaurant". Il existe également le concept "Name" associé à cinq catégories différentes (tels que `"restaurant_name"`, `"movie_name"` et `"location_name"`). Nous avons fusionné les deux en deux étapes : **(1)** la première consiste à fusionner les deux données sans changer les ontologies. **(2)** La deuxième consiste à changer les concepts de SNIPS en une seule entité (`"movie_name"` devient `"name"`) et les intentions en une forme `"Domaine-Acte"` (`"SearchScreeningEvent"` devient `"Event-Search"`). En ce qui concerne MultiWOZ, nous avons étendu les concepts en leur forme de langage naturel (`"Dest"` devient `"Destination"`), nous les avons mis tous en minuscules et nous avons appris un modèle mT5 (Xue *et al.*, 2020) qui prend en entrée les énoncés et génère les représentations structurées comme indiqué dans la figure 3, codées en format Penman.

Les performances globales dans le tableau 9 montrent que la fusion des deux corpora n'a pas engendré de résultats significatifs. Bien qu'une légère baisse ait été observée pour SNIPS au niveau de l'accuracy globale, MultiWOZ n'a pas vraiment changé. Nous avons ensuite examiné les performances au niveau des concepts composés comportant le mot `"name"` dans leur nom et simplifiés au format MultiWOZ.

Le tableau 10 présente les performances des F1-mesures au niveau Intention(slot,valeur) correspondant aux slots modifiés et leurs domaines respectifs.

Nous constatons une amélioration au niveau des résultats pour le slot `"movie_name"` et en particulier pour `"restaurant_name"` suite à la modification de l'ontologie de SNIPS (MS-S^O), avec une augmentation de 16%. Ce label partage en effet le domaine et le nom du concept avec MultiWOZ, et il semble que le modèle génératif a bien capturé la sémantique derrière.

En revanche, les performances pour `"location_name"` ont diminué. Les résultats pour `"object_name"` ont également diminué avec l'intention `"SearchScreeningEvent"` mais ils s'améliorent avec 1% avec l'intention `"RateBook"`. Tout cela montre que l'association entre les différents labels affectent leurs significations, et il est possible pour les modèles génératifs de les prédire si on peut les représenter d'une manière cohérente. L'augmentation significative de certains concepts met en évidence l'importance de l'unification des ontologies. En effet, lorsque les énoncés

et leurs labels partagent une sémantique commune, et sont bien définis et unifiés sous le même label, cela peut contribuer à renforcer les performances des systèmes N/SLU.

Nous envisageons d'approfondir cette approche et étudier précisément les ontologies en exploitant à la fois les représentations structurées des frames sémantiques et le potentiel des modèles génératifs pour l'unification des jeux de données en compréhension du langage naturel et parlé.

SNIPS			
	S-S	**MS-S**	**MS-S^O**
F1 intention	98,1	97,8	98,6
F1 (concept,valeur)	95,0	94,8	94.9
F1 Intent(concept,valeur)	94,7	94,6	94,6
Accuracy global	88,7	87,8	88,0
MultiWOZ 2.3			
	M-M	**MS-M**	**MS-M^O**
F1 intention	96,2	96,2	96,3
F1 (concept,valeur)	94,7	94,9	94,9
F1 Intent(concept,valeur)	94,1	94,2	94,3
Accuracy global	87,6	87,7	87,5

TABLE 9 – Résultats des expériences sur la sémantique des labels : apprentissage et test sur SNIPS (**S-S**), apprentissage et test sur MultiWOZ (**M-M**), apprentissage sur MultiWOZ et SNIPS sans modifications de l'ontologie et test sur SNIPS **(MS-S**), apprentissage sur MultiWOZ et SNIPS sans modifications de l'ontologie et test sur MultiWOZ (**MS-M**), apprentissage sur MultiWOZ et SNIPS avec modifications de l'ontologie et test sur SNIPS (**MS-S**^{O}), apprentissage sur MultiWOZ et SNIPS avec modifications de l'ontologie et test sur MultiWOZ (**MS-M**^{O})

	S-S	**MS-S**	**MS-S^O**
`SearchScreeningEvent+movie_name (event-search+name)`	86,7	77,1	89,6
`SearchScreeningEvent+location_name (event-search+name)`	97,9	97,9	91,7
`SearchCreativeWork+object_name (work-search+name)`	85,9	85,6	82,9
`AddToPlaylist+entity_name (music-add+name)`	74,6	80,0	71,9
`RateBook+object_name (book-rate+name)`	95,0	97,5	96,3
`BookRestaurant+restaurant_name (restaurant-book+name)`	73,3	83,9	89,7

TABLE 10 – F1 mesure niveau Inent(concept,valeur) pour les concepts composés par le slot "name" : format d'origine (format modifié)

4 Conclusion

La compréhension du langage naturel et parlé est une tâche fondamentale dans les systèmes des dialogues. Les deux tâches connues visent à comprendre les commandes de l'utilisateur et ses interactions avec un agent robotique. Dans cet article, nous avons présenté les différents jeux de données utilisés dans ce domaine et nous avons comparé leurs ontologies et leurs schémas d'annotation. Les représentations sémantiques structurées et la méthode de graphe ont été mises en valeur pour leur potentiel à refléter la hiérarchie sémantique entre les frames sémantiques et à exploiter le contexte. La sémantique des labels a également fait l'objet d'une discussion dans nos projets de recherche basés fondamentalement sur des expériences sur le corpus conversationnel MultiWOZ. Dans le but d'unifier

les données et d'exploiter mieux le contexte pour construire des systèmes robustes, nous envisageons d'approfondir nos études des ontologies et des représentations structurées tout en exploitant l'histoire conversationnelle.

Références

ABROUGUI R., DAMNATI G., HEINECKE J. & BÉCHET F. (2022). Étiquetage ou génération de séquences pour la compréhension automatique du langage en contexte d'interaction ? In *Traitement Automatique des Langues Naturelles (TALN 2022)*, p. 64–73 : ATALA.

AGHAJANYAN A., MAILLARD J., SHRIVASTAVA A., DIEDRICK K., HAEGER M., LI H., MEHDAD Y., STOYANOV V., KUMAR A., LEWIS M. *et al.* (2020). Conversational semantic parsing. *arXiv preprint arXiv :2009.13655*.

ASRI L. E., SCHULZ H., SHARMA S., ZUMER J., HARRIS J., FINE E., MEHROTRA R. & SULEMAN K. (2017). Frames : a corpus for adding memory to goal-oriented dialogue systems. *arXiv preprint arXiv :1704.00057*.

ATHIWARATKUN B., SANTOS C. N. D., KRONE J. & XIANG B. (2020). Augmented natural language for generative sequence labeling. *arXiv preprint arXiv :2009.13272*.

BAKER C. F., FILLMORE C. J. & LOWE J. B. (1998). The berkeley framenet project. In *COLING 1998 Volume 1 : The 17th International Conference on Computational Linguistics*.

BANARESCU L., BONIAL C., CAI S., GEORGESCU M., GRIFFITT K., HERMJAKOB U., KNIGHT K., KOEHN P., PALMER M. & SCHNEIDER N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, p. 178–186.

BASTIANELLI E., VANZO A., SWIETOJANSKI P. & RIESER V. (2020). Slurp : A spoken language understanding resource package. *arXiv preprint arXiv :2011.13205*.

BÉCHET F. & RAYMOND C. (2018). Is atis too shallow to go deeper for benchmarking spoken language understanding models ? In *InterSpeech 2018*, p. 1–5.

BÉCHET F. & RAYMOND C. (2019). Benchmarking benchmarks : introducing new automatic indicators for benchmarking spoken language understanding corpora. In *Interspeech*.

BECHET F., RAYMOND C., HAMANE A., ABROUGUI R., MARZINOTTO G. & DAMNATI G. (2022). Can we predict how challenging spoken language understanding corpora are across sources, languages, and domains ? In *Conversational AI for Natural Human-Centric Interaction : 12th International Workshop on Spoken Dialogue System Technology, IWSDS 2021, Singapore*, p. 33–45 : Springer.

BELLOMARIA V., CASTELLUCCI G., FAVALLI A. & ROMAGNOLI R. (2019). Almawave-slu : A new dataset for slu in italian. *arXiv preprint arXiv :1907.07526*.

BONIAL C., DONATELLI L., ABRAMS M., LUKIN S., TRATZ S., MARGE M., ARTSTEIN R., TRAUM D. & VOSS C. (2020). Dialogue-amr : abstract meaning representation for dialogue. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, p. 684–695.

BUDZIANOWSKI P., WEN T.-H., TSENG B.-H., CASANUEVA I., ULTES S., RAMADAN O. & GAŠIĆ M. (2018). Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *arXiv preprint arXiv :1810.00278*.

CHEN X., GHOSHAL A., MEHDAD Y., ZETTLEMOYER L. & GUPTA S. (2020). Low-resource domain adaptation for compositional task-oriented semantic parsing. *arXiv preprint arXiv :2010.03546*.

CHENG J., AGRAWAL D., ALONSO H. M., BHARGAVA S., DRIESEN J., FLEGO F., GHOSH S., KAPLAN D., KARTSAKLIS D., LI L. *et al.* (2020). Conversational semantic parsing for dialog state tracking. *arXiv preprint arXiv :2010.12770*.

COOPE S., FARGHLY T., GERZ D., VULIĆ I. & HENDERSON M. (2020). Span-convert : Few-shot span extraction for dialog with pretrained conversational representations. *arXiv preprint arXiv :2005.08866*.

COUCKE A., SAADE A., BALL A., BLUCHE T., CAULIER A., LEROY D., DOUMOURO C., GISSELBRECHT T., CALTAGIRONE F., LAVRIL T. *et al.* (2018). Snips voice platform : an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv :1805.10190*.

DAHL D. A., BATES M., BROWN M. K., FISHER W. M., HUNICKE-SMITH K., PALLETT D. S., PAO C., RUDNICKY A. & SHRIBERG E. (1994). Expanding the scope of the atis task : The atis-3 corpus. In *Human Language Technology : Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*.

DESOT T., RAIMONDO S., MISHAKOVA A., PORTET F. & VACHER M. (2018). Towards a french smart-home voice command corpus : Design and nlu experiments. In *Text, Speech, and Dialogue : 21st International Conference, TSD 2018, Brno, Czech Republic, September 11-14, 2018, Proceedings 21*, p. 509–517 : Springer.

DEVILLERS L., MAYNARD H., ROSSET S., PAROUBEK P., MCTAIT K., MOSTEFA D., CHOUKRI K., CHARNAY L., BOUSQUET C., VIGOUROUX N. *et al.* (2004). The french media/evalda project : the evaluation of the understanding capability of spoken language dialogue systems. In *LREC*.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2018). Bert : Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv :1810.04805*.

DINARELLI M. (2010). *Spoken language understanding : from spoken utterances to semantic structures*. Thèse de doctorat, University of Trento.

DINARELLI M., QUARTERONI S., TONELLI S., MOSCHITTI A. & RICCARDI G. (2009). Annotating spoken dialogs : from speech segments to dialog acts and frame semantics. In *Proceedings of SRSL 2009, the 2nd Workshop on Semantic Representation of Spoken Language*, p. 34–41.

ERIC M., GOEL R., PAUL S., SETHI A., AGARWAL S., GAO S. & HAKKANI-TUR D. (2019). Multiwoz 2.1 : Multi-domain dialogue state corrections and state tracking baselines. *arXiv :1907.01669*.

FITZGERALD J., HENCH C., PERIS C., MACKIE S., ROTTMANN K., SANCHEZ A., NASH A., URBACH L., KAKARALA V., SINGH R. *et al.* (2022). Massive : A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages. *arXiv preprint arXiv :2204.08582*.

GUPTA S., SHAH R., MOHIT M., KUMAR A. & LEWIS M. (2018). Semantic parsing for task oriented dialog using hierarchical representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 2787–2792, Brussels, Belgium : Association for Computational Linguistics. DOI : 10.18653/v1/D18-1300.

HAN T., LIU X., TAKANOBU R., LIAN Y., HUANG C., WAN D., PENG W. & HUANG M. (2020). Multiwoz 2.3 : A multi-domain task-oriented dialogue dataset enhanced with annotation corrections and co-reference annotation. *arXiv :2010.05594*.

HEMPHILL C. T., GODFREY J. J. & DODDINGTON G. R. (1990). The atis spoken language systems pilot corpus. In *Speech and Natural Language : Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*.

HU X., DAI J., YAN H., ZHANG Y., GUO Q., QIU X. & ZHANG Z. (2022). Dialogue meaning representation for task-oriented dialogue systems. *arXiv preprint arXiv :2204.10989*.

KASPER R. T. (1989). A flexible interface for linking applications to penman's sentence generator. In *Speech and Natural Language : Proceedings of a Workshop Held at Philadelphia, Pennsylvania, February 21-23, 1989*.

LEE S., ZHU Q., TAKANOBU R., LI X., ZHANG Y., ZHANG Z., LI J., PENG B., LI X., HUANG M. *et al.* (2019). Convlab : Multi-domain end-to-end dialog system platform. *arXiv preprint arXiv :1904.08637*.

LEFEVRE F., MOSTEFA D., BESACIER L., QUIGNARD M., CAMELIN N., FAVRE B., JABAIAN B., BARAHONA L. M. R. *et al.* (2012). Leveraging study of robustness and portability of spoken language understanding systems across languages and domains : the portmedia corpora. In *The International Conference on Language Resources and Evaluation*.

LI H., ARORA A., CHEN S., GUPTA A., GUPTA S. & MEHDAD Y. (2020). Mtop : A comprehensive multilingual task-oriented semantic parsing benchmark. *arXiv preprint arXiv :2008.09335*.

LOOS B. (2006). Scaling natural language understanding via user-driven ontology learning. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, p. 33–40.

MA Y., AUDIBERT L. & NAZARENKO A. (2009). Ontologies étendues pour l'annotation sémantique. In *20es Journées Francophones d'Ingénierie des Connaissances*, p. 205–216.

PESKOV D., CLARKE N., KRONE J., FODOR B., ZHANG Y., YOUSSEF A. & DIAB M. (2019). Multi-domain goal-oriented dialogues (multidogo) : Strategies toward curating and annotating large scale dialogue data. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, p. 4526–4536.

PORTET F., CAFFIAU S., RINGEVAL F., VACHER M., BONNEFOND N., ROSSATO S., LECOUTEUX B. & DESOT T. (2019). Context-aware voice-based interaction in smart home-vocadom@ a4h corpus collection and empirical assessment of its usefulness. In *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, p. 811–818 : IEEE.

QIN L., XIE T., CHE W. & LIU T. (2021). A survey on spoken language understanding : Recent advances and new frontiers. *arXiv preprint arXiv :2103.03095*.

SCHUSTER S., GUPTA S., SHAH R. & LEWIS M. (2019). Cross-lingual transfer learning for multilingual task oriented dialog. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 3795–3805, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : 10.18653/v1/N19-1380.

SHAH P., HAKKANI-TÜR D., TÜR G., RASTOGI A., BAPNA A., NAYAK N. & HECK L. (2018). Building a conversational agent overnight with dialogue self-play. *arXiv preprint arXiv :1801.04871*.

TOMASELLO P., SHRIVASTAVA A., LAZAR D., HSU P.-C., LE D., SAGAR A., ELKAHKY A., COPET J., HSU W.-N., ADI Y. *et al.* (2023). Stop : A dataset for spoken task oriented semantic parsing. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, p. 991–998 : IEEE.

TUR G. & DE MORI R. (2011). *Spoken language understanding : Systems for extracting semantic information from speech*. John Wiley & Sons.

WELD H., HUANG X., LONG S., POON J. & HAN S. C. (2022). A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys*, **55**(8), 1–38.

XUE L., CONSTANT N., ROBERTS A., KALE M., AL-RFOU R., SIDDHANT A., BARUA A. & RAFFEL C. (2020). mt5 : A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv :2010.11934*.

YE F., MANOTUMRUKSA J. & YILMAZ E. (2021). Multiwoz 2.4 : A multi-domain task-oriented dialogue dataset with essential annotation corrections to improve state tracking evaluation.

ZANG X., RASTOGI A., SUNKARA S., GUPTA R., ZHANG J. & CHEN J. (2020). Multiwoz 2.2 : A dialogue dataset with additional annotation corrections and state tracking baselines. In *WNLPC AI, ACL'20*.

ZETTLEMOYER L. S. & COLLINS M. (2012). Learning to map sentences to logical form : Structured classification with probabilistic categorial grammars. *arXiv preprint arXiv :1207.1420*.