

Étiquetage ou génération de séquences pour la compréhension automatique du langage en contexte d'interaction?

Rim Abrougui^{1,2} Géraldine Damnati¹ Johannes Heinecke¹ Frédéric Béchet²

(1) Orange Innovation, 22307 Lannion, France

(2) Aix-Marseille Université, 13288 Marseille, France

rim.abrougui@orange.com

RÉSUMÉ

La tâche de compréhension automatique du langage en contexte d'interaction (NLU pour *Natural Language Understanding*) est souvent réduite à la détection d'intentions et de concepts sur des corpus mono-domaines annotés avec une seule intention par énoncé. Afin de dépasser ce paradigme, nous cherchons à aborder des référentiels plus complexes en visant des représentations sémantiques structurées au-delà du simple modèle intention/concept. Nous nous intéressons au corpus MultiWOZ, couramment utilisé pour le suivi de l'état du dialogue. Nous questionnons la projection de ces annotations sémantiques complexes pour le NLU, en comparant plusieurs approches d'étiquetage de séquence, puis en proposant un nouveau formalisme inspiré des méthodes de génération de graphe pour la modélisation sémantique AMR. Nous discutons enfin le potentiel des approches génératives.

ABSTRACT

Sequence tagging or sequence generation for Natural Language Understanding?

Natural Language Understanding (NLU) in an interaction context is often reduced to Intent detection and Slot filling on mono-domain corpora annotated with single Intent utterances. In order to go beyond this paradigm, we seek to approach more complex ontologies towards structured semantic representations beyond the simple Intent / Slot filling model. For this purpose, we are interested in the MultiWOZ dataset, commonly used for Dialog State Tracking. We study different strategies to project these complex semantic annotations for NLU, by comparing several sequence-tagging approaches, then by proposing a new formalism inspired by recent graph generation methods for AMR semantic modeling. Finally, we discuss the potential of generative approaches.

MOTS-CLÉS : Compréhension du langage, étiquetage de séquence, génération seq2seq.

KEYWORDS: Natural Language Understanding, sequence tagging, seq2seq generation.

1 Introduction

La compréhension automatique de la parole (ou SLU/NLU pour *Spoken / Natural Language Understanding*) est une tâche traditionnellement étudiée dans le contexte des systèmes de dialogue humain-machine où chaque tour de parole correspond à une requête ou une réponse de l'utilisateur face à une question du système. Les limites inhérentes à ce type de systèmes ont permis de simplifier les modèles sémantiques utilisés pour représenter le sens de chaque énoncé en exploitant le fait que chaque requête ne devait contenir qu'un seul but et ne faisait pas intervenir le contexte du dialogue pour en expliciter le sens. C'est ainsi que les modèles sémantiques à plat du type *domaine, intention,*

concept+valeur (domain, intent, slot-values) ont été exploités dans des corpus de référence tels que ATIS, SNIPS, M2M ou encore MEDIA. Pour pouvoir prédire de telles représentations, les modèles d'étiquetage de séquences se sont imposés comme étant la solution la plus performante, depuis les HMM, CRF, LSTM (Hakkani-Tür *et al.*, 2016) jusqu'aux modèles de réseaux de neurones (*Deep Neural Network*, DNN) à base d'attention (Goo *et al.*, 2018; Chen *et al.*, 2019), ces derniers modèles atteignant des performances proches de la perfection dans des conditions d'utilisation *benchmark*. Cependant les limites de ces modèles sont doubles : d'une part les modèles d'étiquetage de séquences sont des modèles de classification, ils nécessitent donc de disposer d'un corpus d'apprentissage suffisant pour chaque classe à prédire, tout changement de domaine ou d'ontologie verra les performances du modèle s'écrouler, ne permettant pas de mutualiser les corpus d'applications ou de schémas d'annotation différents ; d'autre part les représentations sémantiques à *plat* représentent difficilement des structures sémantiques plus complexes dans lesquelles des énoncés peuvent contenir plusieurs intentions et intégrer du contexte, comme par exemple dans le corpus MultiWOZ.

Afin de s'attaquer à ces deux limitations, cet article propose deux contributions. La première est une comparaison, sur le corpus MultiWOZ, de deux paradigmes de NLU/SLU pour la prédiction de structures sémantiques : l'étiquetage de séquences et la prédiction directe d'un graphe sémantique par une approche générative de type séquence-à-séquence. La deuxième contribution concerne une approche originale de type séquence-à-séquence, en proposant un schéma d'encodage de graphes sémantiques de NLU/SLU inspiré du formalisme de Penman (Kasper, 1989) utilisé pour les représentations sémantiques AMR (*Abstract Meaning Representation*) (Goodman, 2020) qui permet de s'abstraire d'un modèle d'annotation développé pour un domaine particulier, ouvrant la porte à une plus grande généralité des modèles NLU/SLU.

2 Le corpus MultiWOZ

MultiWOZ est un jeu de données multi-domaines à grande échelle, en langue anglaise, fréquemment utilisé dans le suivi de l'état du dialogue (*Dialog State Tracking*), la politique de dialogue et les tâches de génération de dialogue. Il a été collecté suivant l'approche magicien d'Oz par *crowd-sourcing*. La première version de MultiWOZ a été publiée en 2018 (Budzianowski *et al.*, 2018) pour encourager les progrès dans la construction de systèmes de dialogue en utilisant une approche supervisée. Il est constitué d'un ensemble de dialogues pour lesquels chaque énoncé (système et utilisateur) est annoté par une séquence d'actes de dialogue dont la définition a pu varier avec le temps. La première version de ce corpus était très bruitée avec par exemple certaines valeurs annotées plusieurs fois par le même acte de dialogue, des actes de dialogue manquants, des valeurs attribuées à de faux concepts, etc. De ce fait, plusieurs versions du corpus ont été produites (*MultiWOZ2.1* (Eric *et al.*, 2019), *MultiWOZ2.2* (Zang *et al.*, 2020), *MultiWOZ2.3* ou *MultiWOZ-coref* (Han *et al.*, 2020) et *MultiWOZ2.4* (Ye *et al.*, 2021)) avec pour effet de corriger certaines erreurs d'annotation et de simplifier et rationaliser le format des annotations. Nous avons choisi de travailler sur la version 2.3 qui s'avère être la plus stable niveau des actes de dialogue et de la tokenization des énoncés, même si elle n'est pas complètement exempte d'erreurs d'annotation.

L'annotation sémantique du corpus MultiWOZ est produite sous la forme de cadres sémantiques (ou *Frames* sémantiques)¹. Un cadre sémantique est composé d'une intention avec un ensemble d'arguments sous la forme d'une paire concept/valeur (ou *slot/value*). L'intention est elle-même une combinaison d'un domaine et d'un acte de dialogue. L'annotation peut être accompagnée d'une

1. L'annotation des cadres sémantiques correspond à l'attribut "dialog_act" dans les fichiers d'annotation au format *json* mais nous préférons conserver la terminologie classiquement utilisée en NLU/SLU

indication sur l’empan des valeurs pour les concepts ayant des valeurs ouvertes. Dans le premier exemple du tableau 1, l’intention est constituée du domaine `Hotel` et de l’acte `Inform` et est associée aux concepts `Parking` et `Price`. La valeur `yes` n’a pas de correspondance exacte dans le texte, alors que l’empan est fourni pour la deuxième valeur (7, 7). Deux intentions différentes sont présentes dans le deuxième exemple, ainsi que dans le troisième où ils concernent de surcroît deux domaines différents et avec des supports identiques. Enfin le quatrième exemple illustre le fait que les annotations sont intrinsèquement contextuelles. En effet, l’association entre *"the postcode for that"* et le domaine `Attraction` ne peut pas être résolue en ne tenant compte que de l’énoncé courant. Nous ne tiendrons pas compte du contexte dans cette étude mais l’aborderons dans de futurs travaux. Le corpus MultiWOZ 2.3 contient 8 domaines applicatifs couvrant *Train, Taxi, Hotel*,

énoncés	annotation en acte de dialogues
◊ I need to make sure it’s cheap and I need parking	"Hotel-Inform": [["Parking", "yes"], ["Price", "cheap", 7, 7]]
◊ I’ll need the adress for one that does have wifi please	"Hotel-Inform": [["Internet", "yes"]], "Hotel-Request": [["Addr", "?"]]
◊ The taxi should depart from parkside police station	"Police-Inform": [["Name", "parkside police station", 6, 8]], "Taxi-Inform": [["Depart", "parkside police station", 6, 8]]
◊ Can I get the postcode for that and I also need to book a taxi to the Golden Wok	"Taxi-Inform": [["Dest", "Golden Wok", 17, 18]], "Attraction-Request": [["Post", "?"]]

TABLE 1 – Exemples d’énoncés annotés dans le corpus MultiWOZ2.3

Restaurant, Attraction, Hospital, Bus et *Police* auxquels s’ajoute le domaine *General* pour les actes de dialogue génériques de l’interaction comme *Bye, Greet, Thank* et *Welcome*. Les valeurs normalisées sans empan associé, aussi appelées *categorical slots* dans la littérature, sont `yes`, `no`, `dontcare` et `?` (le dernier correspondant aux questions posées par l’utilisateur et qui est associé souvent à l’acte `Request`) ainsi que `cheap`, `expensive`, `free` et `moderate` qui peuvent être soit associées à un empan si le terme est explicitement prononcé, soit non quand elles sont exprimées par une périphrase (e.g « *I don’t want to pay for wi-fi* » pour la valeur `free`). Les chiffres peuvent également représenter des quantités implicites (par exemple « *for myself* » pour 1 personne). Enfin, la valeur `none`, peut être associée au concept `None` lorsque l’intention n’est pas instancié par un concept ("General-Thank": ["none", "none"]) ou bien à un concept sans valeur associée. Au total le corpus est constitué de 10438 dialogues pour 71521 énoncés utilisateur. Il est annoté avec 27 concepts différents auxquels s’ajoute le concept `none`, 32 intentions de type `Domain-Act` et la combinatoire `Domain-Act+Concept` conduit à 131 éléments.

3 Projection sémantique pour l’étiquetage de séquence

L’approche la plus répandue ces dernières années pour effectuer la tâche de NLU s’appuie sur le paradigme de l’étiquetage de séquence et consiste à projeter les annotations sémantiques dans le formalisme BIO. Les approches jointes visant à prédire simultanément l’intention et les concepts sont abordées, depuis l’avènement des modèles Transformers de type BERT, en associant l’intention au token [CLS] de classification globale et en détectant les concepts sur chaque token concerné avec une étiquette B ou I le cas échéant. Notre objectif principal dans cette section est de projeter l’annotation d’origine du jeu de données MultiWOZ2.3 vers une annotation à plat pour l’étiquetage de séquence. La difficulté ici est de gérer les cas où un énoncé peut avoir plusieurs intentions, où certains concepts ont des valeurs sans pour autant être associés à un empan qui se transposerait en B

ou \mathbb{I} , où un même token peut faire partie d’une valeur pour différents actes de dialogues. La tâche d’étiquetage de séquence sera donc *multi-label* et la projection peut être effectuée de plusieurs façons.

Dans le cadre de la plateforme Convlab (Lee *et al.*, 2019), une projection a été proposée et évaluée avec le modèle MILU (*Multi-Intent Language Understanding*), une extension du modèle multi-tâche OneNet (Kim *et al.*, 2017) pour traiter les actes de dialogue multi-intentions. Convlab-2 (Zhu *et al.*, 2020) a proposé un autre modèle multi-tâche, BERTNLU qui ajoute deux couches MLP (*Multilayer perceptron*) au-dessus de BERT pour prédire respectivement les labels attribués au token [CLS] et les labels BIO attribués aux tokens. Néanmoins, les deux articles ne fournissent pas les détails sur la projection des annotations en actes de dialogue pour constituer ces labels. Une étude du code nous a permis d’en faire une description plus précise. Nous proposons une projection similaire avec cependant quelques différences qui sont illustrées dans le tableau 2. Nous proposons de prédire systématiquement au niveau du token [CLS] l’ensemble des intentions de l’énoncé. Les concepts qui ont un empan associé sont identifiés au niveau des tokens alors que les autres, qui constituent un ensemble fermé, sont entièrement prédits au niveau du [CLS]. Ainsi le premier exemple du tableau 1 est projeté dans le tableau 2 avec deux labels sur le token [CLS], l’un qui porte uniquement l’intention (Hotel-*Inform*) et l’autre qui est une concaténation de l’intention avec le concept et la valeur (Hotel-*Inform*+Parking**yes*). Le concept associé au premier est directement attribué comme label au token *cheap* (B-*Hotel-Inform*+*Price*). La projection de Convlab diffère de la nôtre à deux niveaux : (i) ils n’attribuent au token [CLS] que les intentions associées à des concepts sans empan, (ii) si plusieurs concepts sont annotés sur les mêmes token, comme pour *parkside police station*, ils n’en projettent qu’un seul. Ainsi la projection Convlab n’est multi-label qu’au niveau de la classification globale mais pas au niveau des valeurs en BIO ou l’étiquette choisie est la première du fichier d’annotation.

tokens	notre version	Convlab	tokens	notre version	Convlab
[CLS]	Hotel- <i>Inform</i> Hotel- <i>Inform</i> + Parking* <i>yes</i>	Hotel- <i>Inform</i> + Parking* <i>yes</i>	[CLS]	Police- <i>Inform</i> Taxi- <i>Inform</i>	O
make	O	O	the	O	O
sure	O	O	taxi	O	O
it	O	O	should	O	O
's	O	O	depart	O	O
cheap	B- <i>Hotel-Inform</i> + <i>Price</i>	B- <i>Hotel-</i> <i>Inform</i> + <i>Price</i>	from	O	O
and	O	O	parkside	B- <i>Police-Inform</i> + <i>Name</i> B- <i>Taxi-Inform</i> + <i>Depart</i>	B- <i>Taxi-Inform</i> + <i>Depart</i>
I	O	O	police	I- <i>Police-Inform</i> + <i>Name</i> I- <i>Taxi-Inform</i> + <i>Depart</i>	I- <i>Taxi-Inform</i> + <i>Depart</i>
need	O	O	station	I- <i>Police-Inform</i> + <i>Name</i> I- <i>Taxi-Inform</i> + <i>Depart</i>	I- <i>Taxi-Inform</i> + <i>Depart</i>
parking	O	O			

TABLE 2 – Annotation à plat en BIO

Si le paradigme d’étiquetage de séquence à base de Transformers permet de modéliser conjointement les intentions et les concepts, il n’en demeure pas moins que cette projection à *plat* est intrinsèquement limitée si l’on envisage de prédire une représentation sémantique structurée plus complexe. Nous proposons dans la section suivante une approche générative qui permet d’envisager une représentation sémantique sous forme de graphe pour la modélisation des actes de dialogue.

4 Projection sémantique pour la génération de séquences

4.1 Travaux connexes

Les approches séquence à séquence (*seq2seq*) ont connu récemment un succès grandissant, revisitant les paradigmes de résolution de nombreuses tâches de traitement de la langue. Nous nous intéressons donc à exploiter ce paradigme pour prédire directement les annotations à partir des énoncés sans passer par un modèle d’alignement de type BIO. (Zhao *et al.*, 2021) ont proposé une approche *seq2seq* pour le Dialog State Tracking (DST) en générant directement des séquences de caractères au format `slot1=value1, slot2=value2, . . .` à partir du contexte du dialogue. (Feng *et al.*, 2021) ont pour leur part proposé le modèle *seq2seq-DU* pour la tâche de DST. L’entrée du système est constitué des énoncés et d’une représentation de l’ontologie (*Schema Guided Dialogue*) et l’état de dialogue est généré par une approche à base de pointeurs vers les éléments du schéma et vers les tokens de l’énoncé pour les valeurs associées à des emplacements. Cette approche nécessite ainsi de s’appuyer sur une approche formalisée de l’ontologie du système. Une application au NLU est également proposée en restreignant l’entrée au seul énoncé courant, mais les auteurs ne l’ont évaluée que sur ATIS et SNIPS.

Il faut cependant noter que la tâche de DST et celle de SLU ne sont pas exactement similaires. Dans le cas du DST la tâche consiste à analyser une conversation en utilisant aussi bien les tours de parole du système que ceux de l’utilisateur, l’analyse des énoncés systèmes pouvant désambiguïser l’analyse des énoncés utilisateurs. Ce n’est pas le cas du SLU qui ne peut utiliser la réponse du système pour analyser une requête d’un utilisateur. Nous souhaitons d’une part rester dans un cadre de SLU multi-domaines et multi-intentions et d’autre part, nous souhaitons aller au-delà de la simple génération des annotations telles que décrites au tableau 1. Nous proposons ainsi un formalisme original présentant plusieurs avantages et plus de potentiel. En effet, pour aller plus loin dans la compréhension du langage il semble nécessaire de dépasser le simple paradigme intention/concept en allant vers des représentations de type graphe au plus grand pouvoir expressif. Nous nous sommes inspirés pour cela des travaux récents en analyse sémantique qui ont vu l’émergence de modèles *seq2seq* permettant de prédire des représentations sémantiques au format AMR (*Abstract Meaning Representation* (Banarescu *et al.*, 2013)) avec un important saut qualitatif.

Les graphes sémantiques produits dans le cadre de l’analyse AMR sont transcrits dans le format Penman (Kasper, 1989) qui produit une représentation linéaire d’une structure de graphe à l’aide de parenthèses et de variables pour exprimer des relations et des concepts.

4.2 ARMILU : représentation abstraite en graphe pour la compréhension du langage multi-intention

Nous avons ainsi choisi de transposer l’annotation sémantique de MultiWOZ à l’aide du format Penman pour deux principales raisons. D’une part, nous pouvons nous appuyer sur des outils de manipulation (Goodman, 2020) et de génération de ces graphes mis au point pour AMR qui permettent de garantir l’obtention de graphes formellement valides. D’autre part, ce formalisme ouvre des perspectives vers des annotations sémantiques plus complexes dans la perspective d’aller au-delà des représentations « à plat ». Les outils Spring (Bevilacqua *et al.*, 2021) et sa variante X-AMR (Cai *et al.*, 2021) mis au point pour prédire des graphes AMR s’appuient sur le modèle pré-entraîné mBART (Liu *et al.*, 2020). La librairie AMRlib (<https://github.com/bjascob/amrlib>)

quant à elle s'appuie sur le modèle T5 (Raffel *et al.*, 2020) et nous l'avons modifiée pour pouvoir utiliser la version multilingue mT5 (Xue *et al.*, 2021). Tous deux sont agnostiques de la nomenclature des relations et intègrent un module de validation de la syntaxe Penman. Nous proposons ainsi la transposition suivante illustrée par les figures 2 et 1 pour respectivement un énoncé multi-intentions et un énoncé mono-intention.

« I need train reservation from **norwich** to **cambridge**. »

```
"dialog_act": {"Train-Inform":
  [{"Dest",
    "cambridge", 7, 7],
  ["Depart",
    "norwich", 5, 5]}}
```

```
(Tm1 / Train-Inform
 :Dest (DesTr2 / Dest
 :name (n2 / name
 :op1 "cambridge"))
 :Depart (DepTr4 / Depart
 :name (n2 / name
 :op1 "norwich"))
```

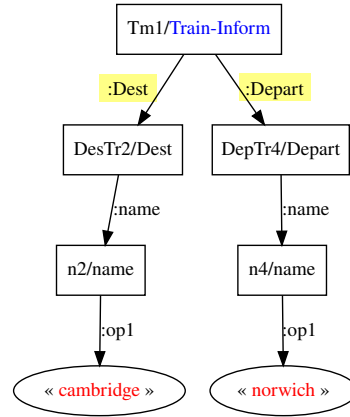


FIGURE 1 – Annotation Penman et visualisation graphique : exemple mono-intention

Pour chaque énoncé utilisateur, nous avons fixé un focus sur l'intention *Domain-Act* dans le cas d'un énoncé comportant une seule intention (figure 1) ou bien sur le symbole *and* qui va déterminer le début de l'annotation en permettant d'avoir plusieurs intentions (figure 2). Dans ce cas, les symboles *:op#* sont utilisés pour marquer les différentes intentions. Les concepts (comme *Dest* et *Post*) sont représentés au même niveau que les rôles sémantiques (*:ARG0*, *:ARG1*, ...) en AMR. Pour les valeurs, si on a une information d'empan, on utilise les symboles *:op#* pour représenter la séquence (en s'inspirant de la représentation des entités nommées pour AMR) et dans ce cas le *Concept* est aussi projeté comme un noeud. Sinon, la valeur est projetée comme un noeud qui a une relation directe avec le *Concept*.

« Can I get the postcode for that? I also need to book a taxi to the **Golden Wok**. »

```
"dialog_act": {"Taxi-Inform":
  [{"Dest",
    "golden wok", 17, 18}],
  "Attraction-Request":
  [{"Post", "?"}]}
```

```
(a / and
 :op1 (Tal / Taxi-Inform
 :Dest (DesTal / Dest
 :name (n1 / name
 :op1 "golden"
 :op2 "wok"))
 :op2 (At2 / Attraction-Request
 :Post (a2 / unknown))
```

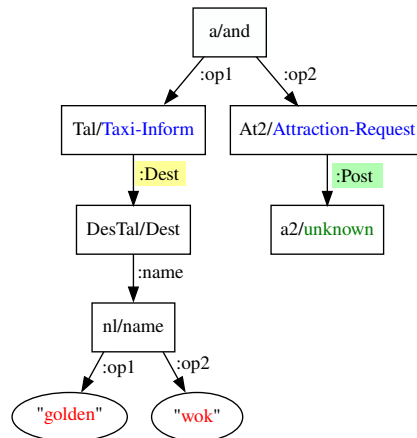


FIGURE 2 – Annotation Penman et visualisation graphique : exemple multi-intention

Nous avons nommé la projection proposée par l'acronyme ARMILU pour *Abstract Representation for Multi-Intent spoken and natural Language Understanding*.

5 Expériences

5.1 Protocole expérimental

Pour les expériences d’étiquetage de séquence nous avons implémenté un *finetuning* de mBERT pour une classification multi-label avec notre projection du tableau 2. Le modèle a été entraîné sur 50 époques au maximum avec une taille de batch égale à 100. Le taux d’apprentissage est égal à $5,0e-05$ et nous avons utilisé la fonction de perte d’entropie croisée binaire. Nous avons évalué le modèle multitâche BERTNLU² avec la projection Convlab décrite également au tableau 2. Nous n’avons pas changé les paramètres par défaut de ce modèle. Le modèle *seq2seq* a été entraîné à l’aide de la librairie AMRlib en utilisant le modèle pré-entraîné mT5 et les paramétrages par défaut de la librairie.

Nous avons respecté le partitionnement (80/10/10) en apprentissage, développement et test du corpus d’origine, la partition de test comportant 7372 énoncés utilisateur.

5.2 Métriques d’évaluation

L’évaluation classique dans les études de NLU se base sur le calcul des performances en termes de F-mesure au niveau intention et concept. Pour les modèles sémantiques acceptant plusieurs intentions par énoncé, une évaluation globale est également réalisée où tous les éléments doivent être reconnus pour considérer un énoncé comme correct. Bien que Convlab-2 (Zhu *et al.*, 2020) propose un outil d’évaluation les notions d’évaluation au niveau intention et concept ne sont pas claires. En effet, leur évaluation dépend de la projection des étiquettes dans la mesure où ils considèrent l’intention comme l’étiquette accolée au symbole [CLS] et les concepts comme les étiquettes accolées aux tokens. Afin de rendre les évaluations comparables indépendamment de la projection choisie, nous proposons une méthode d’évaluation générique. Nous avons choisi pour simplifier d’évaluer directement la prédiction de la liste des éléments *Domaine-Acte(Concept, valeur)*. Ainsi pour le premier énoncé du tableau 1, la référence pour l’évaluation est constituée de la liste {Hotel-Inform(Parking, yes), Hotel-Inform(Price, cheap)}. Les prédictions, qu’elles soient issues des modèles d’étiquetage de séquence ou des modèles *seq2seq* sont également reprojctées dans ce même format. De cette façon, l’évaluation consiste à comparer les listes de références et les listes prédites. Il est toujours possible d’évaluer des F-mesures à différents niveaux ; *Domaine-Acte* seul, (*Concept, valeur*) seul, ou l’élément entier. Nous présentons également une évaluation globale au niveau énoncé où l’ensemble des intentions avec leurs concepts et valeurs associés doivent être correctement reconnus.

5.3 Résultats

	notre modèle BIO	modèle BIO Convlab	ARMILU
<i>Domaine-Acte</i> F1 (R/P)	91,5 (91,6 / 91,5)	91,9 (92,1 / 91,6)	92,0 (90,9 / 93,2)
(<i>Concept, valeur</i>) F1 (R/P)	93,6 (93,6 / 93,6)	93,7 (94,1 / 93,4)	94,1 (92,8 / 95,4)
<i>Domaine-Acte(Concept, valeur)</i> F1 (R / P)	89,6 (89,7 / 89,6)	90,2 (90,5 / 90,0)	89,9 (88,8 / 91,1)
Accuracy globale niveau énoncés	81,0	81,7	82,4

TABLE 3 – Résultats en termes de F1 (Rappel/Précision) à différents niveaux et d’*accuracy* globale au niveau énoncé

2. <https://github.com/thu-coai/ConvLab-2/tree/master/convlab2/nlu/jointBERT/multiwoz>

L'évaluation (c.f. tableau 3) montre que les performances sont proches au niveau des F-mesures des *Domaine-Acte* et (*Concept, valeur*). Au niveau global cependant, le modèle *seq2seq* ARMILU est légèrement meilleur. Ce premier constat est en soi encourageant dans la mesure où nous présentons les premiers résultats de cette approche *seq2seq* sans optimisation particulière. L'évaluation détaillée sur les différents niveaux montre que le modèle de génération de graphe offre une meilleure précision sur les trois premiers niveaux, au détriment du rappel.

5.4 Expériences contrastives

Nous avons créé des partitions de l'ensemble de test selon plusieurs critères liés à des niveaux de difficulté potentiels.

- La première identifie les énoncés qui ont des marques de négation, (*not* ou *n't*), le traitement de la négation étant un facteur de difficulté dans toute analyse sémantique automatique.
- La deuxième a pour but de voir si les modèles ont bien trouvé les valeurs implicite, à savoir les valeurs normalisées sans empan associé (*dontcare, yes, no, none, etc*). Trouver des valeurs implicites sans support précis dans la phrase est plus difficile que d'identifier des entités avec support tels que les entités nommées.
- La troisième concerne les phrases interrogatives (valeur ?).
- La quatrième partition distingue les énoncés qui ont plusieurs labels (soit plusieurs intentions, soit plusieurs concepts, soit un mot pouvant être associé à plusieurs concepts).

Les performances en termes d'accuracy globale sont reportées dans le tableau 4 pour notre modèle BIO et le modèle génératif.

<i>nb énoncés</i>	Négation		Valeurs implicites		Valeur « ? »		Labels	
	avec (470)	sans (6902)	avec (574)	sans (6798)	avec (1499)	sans (5837)	multi (3417)	mono (3667)
modèle BIO	54,9	82,8	50,7	83,5	61,2	86,0	77,4	83,5
modèle ARMILU	60,4	83,9	54,0	84,8	66,3	86,4	78,8	85,0

TABLE 4 – Accuracy globale niveau énoncé sur différentes partitions

Les résultats montrent que le modèle de génération de graphes est meilleur sur toutes les partitions potentiellement difficiles. L'amélioration au niveau des valeurs implicites ou au niveau des énoncés avec des marques de négation confirme que les représentations abstraites et génériques peuvent aider à aller au-delà de la tâche classique de concept/valeur.

6 Conclusion

Nous avons proposé d'envisager le NLU sous l'angle de la génération de graphes sémantiques en nous inspirant du formalisme de Penman utilisé pour la représentation AMR. Avec cette approche qui permet d'avoir une représentation sémantique abstraite, nous obtenons des performances assez proches de l'état de l'art avec un formalisme qui offre de nombreuses perspectives avec la possibilité de s'affranchir des contraintes liées aux projections à plat pour tendre vers des modélisations sémantiques structurées plus complexes.

Références

- 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, Sofia, Bulgaria : Association for Computational Linguistics.
- BEVILACQUA M., BLLOSHMI R. & NAVIGLI R. (2021). One SPRING to Rule Them Both : Symmetric AMR Semantic Parsing and Generation without a Complex Pipeline. In *Proceedings of the AAAI Conference on Artificial Intelligence*, p. 12564–12573.
- 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. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 5016–5026, Brussels, Belgium : Association for Computational Linguistics. DOI : [10.18653/v1/D18-1547](https://doi.org/10.18653/v1/D18-1547).
- CAI D., LI X., CHUN-SING HO J., BING L. & LAM W. (2021). Multilingual AMR Parsing with Noisy Knowledge Distillation. In *Findings of the Association for Computational Linguistics : EMNLP 2021*, p. 2778–2789, Punta Cana, Dominican Republic : Association for Computational Linguistics.
- CHEN Q., ZHUO Z. & WANG W. (2019). Bert for joint intent classification and slot filling. *arXiv preprint arXiv :1902.10909*.
- 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 preprint arXiv :1907.01669*.
- FENG Y., WANG Y. & LI H. (2021). A sequence-to-sequence approach to dialogue state tracking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, p. 1714–1725.
- GOO C.-W., GAO G., HSU Y.-K., HUO C.-L., CHEN T.-C., HSU K.-W. & CHEN Y.-N. (2018). Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 2 (Short Papers)*, p. 753–757.
- GOODMAN M. W. (2020). Penman : An open-source library and tool for AMR graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, p. 312–319, Online : Association for Computational Linguistics.
- HAKKANI-TÜR D., TÜR G., CELIKYILMAZ A., CHEN Y.-N., GAO J., DENG L. & WANG Y.-Y. (2016). Multi-domain joint semantic frame parsing using bi-directional rnn-lstm. In *Interspeech*, p. 715–719.
- 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 preprint arXiv :2010.05594*.
- 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*.

- KIM Y.-B., LEE S. & STRATOS K. (2017). Onenet : Joint domain, intent, slot prediction for spoken language understanding. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, p. 547–553. DOI : [10.1109/ASRU.2017.8268984](https://doi.org/10.1109/ASRU.2017.8268984).
- 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*.
- LIU Y., GU J., GOYAL N., LI X., EDUNOV S., GHAZVININEJAD M., LEWIS M. & ZETTLEMOYER L. (2020). Multilingual denoising pre-training for neural machine translation. *CoRR*, **abs/2001.08210**.
- RAFFEL C., SHAZEER N., LEE K., NARANG S., MATENA M., ZHOU Y., LI W. & J. L. P. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, **21**, 1–67.
- XUE L., CONSTANT N., KALE M., AL-RFOU R., SIDDHANT A., BARUA A. & RAFFEL C. (2021). mT5 : A Massively Multilingual Pre-trained Text-to-Text Transformer. In *Conference of the North American Chapter of the Association for Computational Linguistics*, p. 483–498 : Association for Computational Linguistics.
- 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 *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI, ACL 2020*, p. 109–117.
- ZHAO J., MAHDIEH M., ZHANG Y., CAO Y. & WU Y. (2021). Effective sequence-to-sequence dialogue state tracking. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, p. 7486–7493.
- ZHU Q., ZHANG Z., FANG Y., LI X., TAKANOBU R., LI J., PENG B., GAO J., ZHU X. & HUANG M. (2020). Convlab-2 : An open-source toolkit for building, evaluating, and diagnosing dialogue systems. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.