**Modélisation des entités avec RASA NLU**

**1 - Entités retenues**

Les entités retenues à ce stade pour le développement du chatbot sont:

* **origin** : l’origine du vol
* **destination** : la destination du vol
* **search\_type** : aller simple, alle retour
* **nb\_passengers** : le nombre de passagers
* **discover\_category** : type/motif du voyage, valeurs possibles: *south-america, asia, sun, romantic, visit, etc*

**2 - Rasa NLU Format**

Un fichier d’entrainement pour un modèle avec RASA nlu doit être conforme à une structure (**rasa nlu format**) qui est la suivante:

**====================================**

{  
 "rasa\_nlu\_data": {  
 "common\_examples": [

"text": "je cherche un vol de paris",  
 "intent": "flight\_search",  
 **"entities": [  
 {"start": 21,  
 "end": 26,  
 "value": "paris",  
 "entity": "origin"  
 }  
 ]**

],  
 "regex\_features" : [],  
 "lookup\_tables" : [],  
 "entity\_synonyms": []  
 }  
}

**====================================**

L’outil existant **wit.ai** utilise une représentation différente des entités détectées.

En particulier, **wit** décrit les entités avec les quatre entités cités ci-dessus (start, end, value, entity), auxquels s'ajoutent les attributs **body** pour le texte détecté, **confidence** pour le taux de confiance liés à cette classification.

Par ailleurs, wit propose d’associer des sous-entités à une entité, en introduisant la notion de *rôle*. Par exemple, pour les lieux (villes, endroits, etc.), wit les classe comme appartenant à l’entité *location,* puis attribue un des rôles suivants: *origin, destination*.

Avec la solution qu’on développe, nous n’allons pas retenir cette représentation. Le paragraphe ci-dessous détaille la représentation, c’est-à-dire les valeurs possibles, pour chacune des entités retenues.

**3 - Représentation des entités dans la solution développée**

entity : **origin**

value : ville du texte

entity : **destination**

value : ville du texte

entity : **search\_type\***

value : one\_way / round\_trip

entity : **nb\_passengers\***

value : chiffre 1/2/...

entity : **discover\_category**

value : asia / romantic/ fun/ party/ explore/ cultural/ arts/ family/ america/ etc

**\***: utilisation des **synonymes** pour la détection des entités.