How Utterances & Slot samples affect Intent-matching in Alexa Skills

Chas Sweeting

This is perhaps the biggest area of confusion and uncertainty for developers building Alexa Skills. This article assumes you understand the basics of Alexa utterances and custom slots.

Consider a simple custom Skill with 2 intents:

Consider an Alexa Skill called “My Butler”. The happy path of a simple interaction might be:

> *You: “*Alexa, ask My Butler for juice.”
>
> *Alexa: “*Certainly. Juice on its way. Would you like anything else?”
>
> *You: “*No thank you.”
>
> *Alexa:* “My pleasure.”

You could define the 2 slots (OrderIntent & ThanksIntent) in the Alexa Developer Console as follows:

```
{  "intents": [    {      "slots": [        {          "name": "food_item",
"type": "FOOD_TYPE"        }      ],      "intent": "OrderIntent"    },    {
"intent": "ThanksIntent"    }  ]}
```

where the custom slot type (FOOD_TYPE) has is defined as follows:

Custom Slot Types (Optional)
Custom slot types to be referenced by the Intent Schema and Sample Utterances. For general information about cus

Type	Values
FOOD_TYPE	Sandwich | Fruit | Salad | Water | Juice

and corresponding utterances to detect when the user asks for something (‘for’ or ‘to get me’), or replies in the negative (‘no thanks’, etc):

```
OrderIntent for {food_item}OrderIntent to get me {food_item}ThanksIntent Thank you
very muchThanksIntent No thanks. ThanksIntent No thank you.ThanksIntent No that's
all. Thank you.
```

Now to test:

> *You:* “Alexa, ask My Butler for fruit”
>
> *Alexa:* “Certainly. fruit on its way. Would you like anything else?”
>
> *You:* “No thanks”
>
> *Alexa:* “My pleasure”.

No surprises there.

As you may know, you can even ask for food items which were not explicitly specified in your custom slot type. (see Custom slots are not restricted to the custom slot values.) For example let’s ask for ‘pizza’:

> *You: “*Alexa, ask My Butler to get me pizza.”
>
> *Alexa: “*Certainly. Pizza on its way. Would you like anything else?”
>
> *You: “*No thank you.”
>
> *Alexa:* “My pleasure.”

So far, all good. Works as expected.

Now for the ‘WTH?’ moment

As we saw above, we can ask My Butler for food items using either “*Alexa, ask My Butler x”* or *“Alexa, ask My Butler x”* and you get confirmation that the food is on its way.

Let’s try another:

> *You:* “Alexa, ask My Butler for pizza”
>
> *Alexa:* “My pleasure.”

Huh? What is going on here? To recap the things we’ve tried saying:

```
"Alexa, ask My Butler for fruit"       <-- matched correctly. "Alexa, ask My Butler
to get me pizza" <-- matched correctly. "Alexa, ask My Butler for pizza"       <--
matched correctly.
```

Enter the opaque world of Alexa intent matching.

How Alexa does NOT match Intents

It’s natural to think that the process is as follows:

1. You speak to your Echo device. The wake word is detected and voice is streamed to the Amazon/Alexa cloud as an audio signal.
2. Automatic speech recognition (ASR) converts the audio signal to text; with some form of digital signal processing distinguishing the voice from ambient noise.
3. The text is then matched to an Intent based on the wording of the utterance. This would be relatively straight-forward text matching.
4. The Slots are filled for that utterance based on the slot definitions.

For example, when the user asks My Butler *" for pizza"*, which of the following utterances would appear closest? (I've omitted the intent names for clarity).

```
{food_item}        <-- matches " pizza" ?to get me {food_item}Thank you very muchNo
thanks. No thank you.No that's all. Thank you.
```

You'd think it's the first item. And that's how Google Assistant & api.ai work. However, as we've seen, that's NOT what happens with Alexa.

What IS happening?

Nobody knows for sure outside of the Amazon team. But it seems to be a machine learning algorithm which considers the number of words in the utterance, the utterance word itself, as well as the number of words in each sample slot.

Get Chas Sweeting's stories in your inbox

Join Medium for free to get updates from this writer.

Here are some further observations (placed in code blocks so that you can more clearly compare the defined utterances with what the user says):

The actual word used in the utterance matters.

The following **will not match** *"ask My Butler for pizza"* to the OrderIntent:

```
UTTERANCE DEFINITION: ---------------------for {food_item}WHAT THE USER SAYS: -------
------------for pizza                    <-- will  match :(
```

But change the word 'for' to 'buy' and it **will** match *"ask My Butler buy pizza":*

```
UTTERANCE DEFINITION: ---------------------buy {food_item}WHAT THE USER SAYS: -------
------------buy pizza                    <--  match :)
```

This infers Alexa's using a model based on a larger dataset than the utterances within your custom Alexa Skill's Interaction model. (The word 'buy' is more specific and has more tightly constrained uses than the word 'for').

The number of words used in the utterance matters

As we've shown, the following utterance **will not be matched** to OrderIntent when you say *"ask My Butler for pizza".*

```
UTTERANCE DEFINITION: ---------------------for {food_item}WHAT THE USER SAYS: -------
------------for pizza                    <-- will  match :(
```

But the following WILL match to *"ask My Butler for more pizza"* to the intent:

```
UTTERANCE DEFINITION: ---------------------OrderIntent for more {food_item}WHAT THE
USER SAYS: -------------------for more pizza               <--  match :)
```

From a machine learning perspective, that's understandable. More words matching equals greater confidence. Still, how many words is enough?

The total number of words in the Utterance + Sample Slots matters

That just seems whack at first:

```
UTTERANCE DEFINITION: ---------------------for more {food_item}WHAT THE USER SAYS: --
-----------------for more pizza               <--  match for more bottles of beer
<-- will  match
```

This shows clearly that Alexa is not matching intent just based on your utterance definition.

Why is 'bottles of beer' not being matched to the intent? Because Alexa is comparing 'for more bottles of beer' (a 5-word utterance) against all of your defined utterances:

```
OrderIntent for {food_item}OrderIntent to get me {food_item}ThanksIntent Thank you
very muchThanksIntent No thanks. ThanksIntent No thank you.ThanksIntent No that's
all. Thank you.
```

In our case, all of our slot samples were just 1-word long ('sandwich', 'salad', 'water', etc) so we have effectively only defined utterances with a total of 2 or 4 words for the OrderIntent.

Alexa favours the 5-word intent even though none of the actual words match. No, that doesn't seem logical at all to me but that's why I'm writing this — to save you the pain.

To conclude

In short, it seems that Alexa matches what's said by the user against all possible combinations of the defined utterances and slot samples — and using a machine learning algorithm based on a larger dataset (not defined in your Interaction Model).

This is undoubtedly done by Amazon with the goal of increasing accuracy and to improve the customer experience, as opposed to confuse developers :) However, without some insight into specifically how this happens, it introduces a level of variability and uncertainty which

reminds me of developing websites for the early web browsers… you were never quite sure how things would render.

What to do? Keep your skills simple (a good principle in general given the current constraints of the technology) and if if things get complicated, pass everything through as a LITERAL and do your own, manual NLP. At least you won't be second-guessing a dataset and model you have no visibility of.