# Las palabras y los objetos

Julia Milanese

15/10/2021

# 1   Content

This document is an attempt to present all research on the linguistic and programatic understanding with the aim of being able to represent objects, in code and then in VR, through words. The order at this point is, as words are, arbitrary, only based in the order in which I started to do the research, moved by desire or practicallity, depending on the day. So it is, una deriva.

- Automatic Speech Recognition

    - ASR in VR

- Formal Linguistics

- Unreal Engine and Voice

- NLU for objects

# 2  ASR in VR

There are some demos available in YouTube. One of them uses the Android STT and Unreal Engine: - https://www.artstation.com/ayattache: https://www.youtube.com/watch?v=aF5w5F8eGqs

# 3   Objects

Is there one subjects that interest me especially. Up until here I've been thinking at work and the Coconuts project (as simple as the NLU was) in onl one aspect of the problem of taking language understanding to VR. That aspect was the NLU side, e.i., the understanding of language as it is developing right now, as an understanding of the intent of the user and the extraction of some entities. Once that estructure is understood from the "unestructured" text of the user, we can make matches or searches in our system. Let's see an example from out of 3D and VR: working with a menu in a restaurant with voice ordering. It usually consists of understanding an intent from a utterance (The user wants to buy something from the menu), extract the items the user names (A coffee) and match that item againsta a menu. Then we have some kind of structured item that can be sent through the system to complete an sale. This is quite straightforward but in general has a unexpected charachteristic. If it were the case that the user wants more than one coffee, let's say the user wants three coffees, al the same, the resulting structure would be the same with a small difference, it would indicate in some field that the quantity is equal to 3. And that is not a problem until the user wants to add something to only one of them (One with cream). At that point something is clear, although in the menu we have only one item that matches the request, the user is actually thinking of 3 items. They may all share the exact same charachteristics but these are three distinct objects. This might be solved with some logic on top of the NLU successfully, but anyways, it got me thinking on how me relate the NLU on one side and the "reality" it is refering to. In the Coconuts project we came to a really simple, demo like, solution, where each target Mr. Coconut could visit had a word reference. So it was like a really simple language from Wittgenstein's "Philosophical Investigations": we just told the system "If you are shown this sequence of characters, find a match and move to those coordinates previously fixed". And that was ok as a first exploration but you could quickly see that escalability was a challenge for this approach. And this were only targets in space to where the actor could move, we weren't trying to name things that could be moved, nor things with maybe similar names. This simple tagging system is not ideal. So my question in this research is about that, not what current Natural Language Processing can understand from a users text, but how that language signifies things in a 3D world. As Wittgenstein is at this point my under reading (?), and I mean with this, the

only book I never end and continuously read while I do this research, I'll be constantly finding relations between this subject of research and his work in Philosophical Investigations. There are, at this point, two ideas that revolve in my mind. The first one is the idea of a really simple language where words work as tags, a bit like Coconuts worked. The other one is the idea of life forms and language games, because I need to think that not all the infinite things that can be modeled in 3D have to be tagged, but a small language inside a game. Small, at least, in comparison with the "real life" language.

Ok, but I said that language understood as tags is something problematic and useful at the same time. So I'd try to clarify what is the idea I have now: tags as one on one corresponding names to 3D objects is a non escalable proposal, but, if we twist the meaning a bit and think of words as tags corresponding to parameters maybe, maybe, they can be useful to form the understanding what in a multiplicity of objects the user is talking about.