

**Ph.D. in Information Technology  
Thesis Defense**

**October 31th, 2023  
at 14:30  
Aula BIO 1 - Building 21**

**Daniele OCCHIUTO – XXXII Cycle**

**AURORA: NATURAL LANGUAGE SLEEPS IN AN UNCANNY VALLEY OF AMBIGUOUS SEMANTIC CONSTRUCTIONS**

**Supervisor: Prof.ssa Franca Garzotto**

**Abstract:**

My thesis investigates the conversational interaction between humans and machines. It explores the differences between natural and artificial languages. It highlights the innate human ability to build conceptual representations from the conversation, examines why machines fail to do so and questions whether machines will ever be able to interact in natural language.

I build an abstraction layer to uncover the unapparent structure of the conversation. Under this thesis point of view, every conversation admits a precise structure. I propose the conversation structure as a graph built from the syntactic dependency parsing that also carries the conversation semantics. It is possible to follow paths in the graph connecting the relevant concepts of the conversation together: the conversation structure reveals how the syntactic constructions yield the semantic interpretations.

The goal of my thesis is to build an abstraction layer to characterize any conversation. In my thesis I investigate the bond between syntax and semantics in natural language. I generate an abstraction layer on top of the syntactic-semantic bond encapsulating the concepts from the conversation. The conversation is made up by pre-linguistically available “abstractions” (i.e. entities and relations). I investigate whether machines can understand the “abstractions” interpreted effortlessly by humans.

The abstraction layer I propose defines the foundational characteristics of any conversation and is machine interpretable.

The abstraction layer uncovers some of the reasons that cause a gap in the interaction between humans and machines. I review the extent of the conversational gap in natural language interaction. I propose a parallel with the uncanny valley theory from the robotics field. The uncanny valley expresses the affinity perceived by humans when interacting with robots. This thesis studies similar perceived affinities with conversational agents. I exploit the abstraction layer built from syntax to assess the uncanny valley in conversations.

I inspect the part of speech - POS - tags dependencies and drive semantic relations from specific POS that are configurable as arguments of the service I build. As a result, I obtain a graph that constitutes a layer describing the semantic representations constructed through the conversational interaction. I borrow the notion of blueprint drawing from architecture. Since the abstraction layer represents the characteristics of the conversation, and the term blueprint conveys the specifications of an engineering/architectural plan, it is the fittest to convey the conversation plan. The blueprint emphasizes the nature of the layer of abstraction I build: an accurate, reproducible and technical specification of the conversation plan or the conversation structure.

Since my blueprint of the conversation is a graph, I call the level of abstraction proposed by this thesis with the term graphprint. The graph-print is the main output of my work. I am able to analyze the conversation structure by means of graph theory. The graphprint encompasses both syntax and semantics of the natural language in one single abstraction layer. In summary, I capture the structure of a conversation into an interpretable level of

abstraction: a graph.

My thesis proposes the following contributions:

1. The graphprint generation: a conversational blueprint representing the conversation semantics as it arises from syntactic dependencies.
2. The graphprint interpretation: it uncovers a layer that captures one possible origin for the uncanny valley in conversations.
3. The graphprint replication: it is deterministic, reproducible and inspectable by means of known graph algorithms.

In conclusion, my thesis provides a tool that is configurable and useful to generate alternative representation of conversations under one point of view that depends both on syntax and semantics. This suggests a reconciliation between competence and performance explaining the rules of the language and the use of the language respectively. Both mechanisms should have evolved as a package: the construction of the graphprint provides a hint advocating for a symbiotic evolution of syntax together with semantics. The graphprint offers a deterministic approach to portray the conversation plan and reveals which plans signal the appearance of the uncanny valley in conversations.

## **PhD Committee**

Prof. **Pietro Pinoli**, Politecnico di Milano

Prof. **Fabio Tamburini**, Università di Bologna

Prof. **Paolo Gasti**, New York Institute of Technology