# Looking Back at Daisie:
# A Retrospective View on Situated Dialogue Systems Development

**Robert J. Ross**

School of Computing
Dublin Institute of Technology
Ireland
robert.ross@dit.ie

**Abstract**

In this talk we give a brief retrospective analysis of the creation and application of the Daisie architecture for Situated Dialogue that was developed between 2004 and 2009 in Bremen, Germany. In addition to reviewing the design rational and features of Daisie, the chief contributions of this retrospective are a detailing of lessons learned from Daisie's development and use, as well as a proposed agenda for continued work in situated dialogue based on these experiences.

## 1. Introduction

Situated Dialogue Systems pull together research and development in kinematics, linguistics, and sensor systems. As such they arguably represent the pinnacle of both computational linguistics and systems integration research. However while the development of even simple non-situated dialogue applications is notoriously difficult, the creation of complete Situated Dialogue Systems is a highly challenging endeavor that pushes theoretical and practical skills to their limits. Given these challenges, it is very important that we learn from our mistakes and build from our successes wherever possible. Unfortunately however, while many scholarly publications report on advances in Situated Dialogue Systems and review the necessary controlled studies and user evaluations required to substantiate individual Situated Dialogue Systems components, the experience of situated dialogue systems development is often lost in the glow of positive quantitative results.

Given the above, in this talk we will take a step back from individual results and provide a retrospective analysis of the successes and failings of one particular Situated Dialogue System architecture five years after its initial development phase ended. The architecture in question is *Daisie*, which was developed at the Spatial Cognition Research Center in Bremen, Germany between 2004 and 2009 (Ross, 2009). Daisie was developed to provide a research platform for interaction with robots and virtual agents with a high level of explicit representations of discourse structure, agency, and perceptual models. As such we believe our experiences with Daisie may be of direct benefit to other researchers at the intersection of Action, Perception, and Language. In the following we briefly expand on Daisie's features before discussing lessons learned and potential future directions.

## 2. Daisie Architecture

The Daisie architecture was a modular situated dialogue systems integration architecture that included default components for situated dialogue processing including: Speech Recognition, Language Parsing, Language Abstraction, Intention Management, and Language Integration. The Daisie architecture, illustrated in Figure 1, was chiefly implemented in Java and was designed as a modular architecture which could be customized to specific situated applications through the instantiation of appropriate resources.

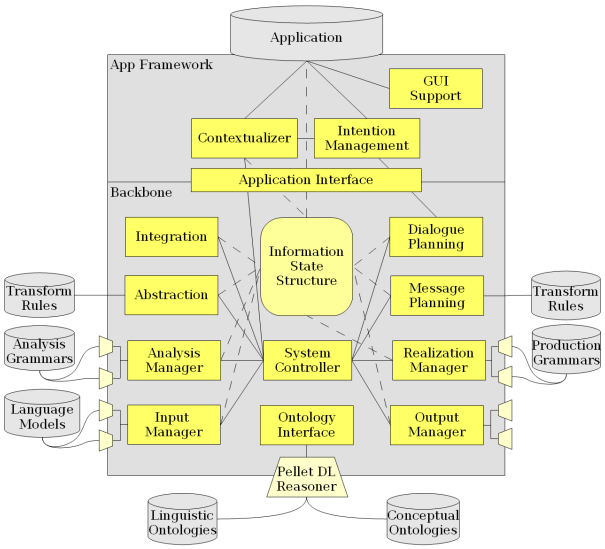


Figure 1: Illustration of the Daisie Architecture

With respect to situated interaction Daisie's main design features included:

- **Intention Management System** - Daisie supported an intention management system that could be extended to provide explicit intention support in given situated dialogue applications. This work was key to managing dialogue regarding ongoing and planned actions. The Intention Management system was designed to be separate from but compatible with a traditional Information State representation.
- **Two-Level Semantics** - Daisie was explicitly designed with the assumption that a multi-ontology or minimally a two-level semantics model was required in capturing linguistic and conceptual knowledge. Here linguistic semantics as organized by the Generalized Upper Model (Bateman et al., 2010) were used to

capture the surface form of language at the grammar interface to language parsers and text generators. On the other hand application knowledge - i.e., the world and actions models of a given situated agent - were assumed to be otherwise structured and an important aspect of the Daisie architecture was producing suitable mappings between these representation layers.

- **Language Contextualization Model** - Daisie provided an explicit language contextualization model for attempting to determine the 'relevance' of any given user utterance to current situational and intention representations.
- **Modular Dialogue Manager Design** - Daisie supported both Information State and Graph Based Dialogue management through a pluggable architecture.
- **Linguistic Alignment Modeling** - Through the shared information state Daisie supported the shaping of linguistic choices to align where possible and appropriate with the user. Alignment here referred to not only lexical and syntactic alignment, but also alignment of higher level situated features such as spatial frames of reference.

## 3. Lessons Learned

In developing and applying the Situated Dialogue Architecture a number of lessons were learned that may be of benefit to those working in this domain. Here we consider a number of these.

### 3.1 The Grammar Process

Whereas keyword spotting is often adequate for question answering systems as well as non-situated dialogue applications, the need to precisely describe spatial relations, subject object relations, and temporal constraints requires a high level of grammar sophistication. These grammars are required both for language production and analysis and represent one of the greatest localization requirements for any situated dialogue application.

Unfortunately wide coverage grammars which provide sophisticated semantics remain elusive. In Daisie, like other contemporary situated dialogue applications, a decision was made to hand-craft CCG and functional grammars which could be applied for language analysis and production respectively. While the resulting grammars have potential for re-use across other projects - situated dialogue or otherwise - the hand crafting process is ultimately a resource sink which undermined other more interesting work on situated interaction. Situated dialogue application developers should therefore invest enough time in initial resource appraisal in order to reduce the chance of wasted effort on grammar engineering.

### 3.2 The Semantics Interface

For Daisie the hand-crafting of new grammatical resources was motivated chiefly by the desire to apply a specific semantics interface to the grammar. The semantics interface in this case was a linguistic ontology named the Generalized Upper Model Version 3 (Bateman et al., 2010). Since a goal of our research project was the development of this linguistic ontology, this design decision was in context just. However from an external perspective, the development of yet another linguistic ontology or semantic specification of word meaning should be avoided in lieu of pre-existing definitions.

The Generalized Upper Model did provide one suitable linguistic ontology, but even that suffered from its own hand-crafted creation and synthesis. One important lesson learned from this work was that there remains scope for an extension or alternative to the Generalized Upper model that is derived automatically from linguistic data rather than linguistic intuition and examples.

### 3.3 The Spatial Dimension

By definition Situated Dialogue Systems are located in a real or virtual world in which they can act and perceive. While we can build upon existing frameworks for rational agency to provide logics and representations of action that are easily customizable for a specific application, the spatial embedding of situated dialogue provides a far more significant research challenge.

Daisie abstracted away from the specifics of a given spatially situated agent and left such considerations as application specific. While there are obvious advantages in doing so, this belies the fact that how we understand and produce space in a situated context is at the very heart of what it is to be a situated agent. Indeed, different models of spatial representation give rise to different accounts of language production and analysis. Thus while abstracting spatial issues outside of the core Daisie implementation may be pragmatic, it fails to solve the most fundamental re-use question in situated dialogue systems.

## 4. Future Directions

Situated Dialogue Systems development is no doubt challenging, but it need not be a form of academic suicide. The overarching lesson learned from Daisie's development is that projects should not attempt to build complete dialogue systems from the ground up unless of course sufficient resourcing is available.

There is little doubt that situated dialogue systems research would benefit from the creation of a test platform which captures a minimal baseline model of physical capabilities including actions and perceptual modeling. While funded pan-national projects have worked on potential solutions to this need, the community of action, perception and language could do much by simply proposing such a standard on which true innovation could be based.

## References

John A. Bateman, Joana Hois, Robert Ross, and Thora Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174s:10271071.

Robert J. Ross. 2009. *Situated Dialogue Systems: Agency & Spatial Meaning in Task-Oriented Dialogue*. Ph.D. thesis, Universitat Bremen.