Incremental Generation of Visually Grounded Language in Situated Dialogue (demonstration system)

Citation for published version:

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the 9th International Natural Language Generation conference

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
We present a multi-modal dialogue system for interactive learning of perceptually grounded word meanings from a human tutor (Yu et al., ). The system integrates an incremental, semantic, and bidirectional grammar framework – Dynamic Syntax and Type Theory with Records (DS-TTR\(^1\), (Eshghi et al., 2012; Kempson et al., 2001)) – with a set of visual classifiers that are learned throughout the interaction and which ground the semantic/contextual representations that it produces (c.f. Kennington & Schlangen (2015) where words, rather than semantic atoms, are grounded in visual classifiers). Our approach extends Dobnik et al. (2012) in integrating perception (vision in this case) and language within a single formal system: Type Theory with Records (TTR (Cooper, 2005)). The combination of deep semantic representations in TTR with an incremental grammar (Dynamic Syntax) allows for complex multi-turn dialogues to be parsed and generated (Eshghi et al., 2015). These include clarification interaction, corrections, ellipsis and utterance continuations (see e.g. the dialogue in Fig. 1).

Incremental Generation in Context: Generation (surface realisation) in DS-TTR follows exactly the same dynamics as parsing except for an additional subsumption check after every word against some goal concept/context (Purver et al., 2014). Generation is therefore just as incremental and contextual as parsing (Eshghi et al., 2015). This allows for the generation of acceptances, elliptical utterances, short answers, and corrections, as well as continuations. Here, it is the dialogue manager that constructs the goal concept from the semantic analysis of the visual scene, and sends it the grammar for surface realisation – whether this is the semantics of a question, an answer, or an object description (see the system responses in Fig. 1).

\(^1\)Downloadable from: http://sourceforge.net/projects/dylan/
We will show an interactive demonstration of this system at the conference, illustrating how questions, answers and object descriptions are derived and generated incrementally in real-time (Yu et al., ). Work in progress addresses: (1) more complex dialogues; (2) data-driven, incremental dialogue management at the lexical level; (3) integrating the existing DS-TTR model of incremental definite reference generation within the implemented system.

References


