

# Agent Script Generation using Descriptive Text Documents

Tim Balint, Yotam Gingold, and Jan M. Allbeck\*  
George Mason University, 4400 University Drive, MSN 4A5  
Fairfax, VA 22030

**Keywords:** Virtual Humans; Animation Systems; Action From Text

When designing games, artists exert large efforts to create visually compelling scenes. Work such as WordsEye [Coyne and Sproat 2001] can assist artists by parsing natural language texts into static scenes. Complementary to this endeavour is the population of that environment by simulation authors. Adding agents to an environment with plausible behaviors is a time consuming process, as most require individual scripts to control their behavior. This generally degrades variability, as scripts are re-used. In order to assist in creating commands and scenes for virtual actors, we propose a method that can create scripts for agents to plausibly act within a virtual environment. This work is inspired by [Ma 2006], which provides an action to a virtual agent from a single sentence. However, our method works for several agents over longer periods of time.

Natural Language commands to virtual agents are inherently ambiguous and require several modifications in order to properly command virtual actors [Förger et al. 2013]. We constrain the problem domain to remove ambiguities, namely, by not allowing pronouns and requiring all text to be in the present tense. The first allows our system to easily parse out actors and objects in the scene, and keeps users from writing ambiguous sentences by mistake, hopefully making it easier for a novice user to debug scripts. The second constraint comes naturally from the problem domain, where virtual agents are performing actions to display to a user. We also define three types of actions that an agent can perform: postural, event, and gestural actions. Postural actions control the end state of an agent, event actions update the state of the world, and gestural actions are atomic actions with no effect on the world state.

One challenge of using natural language to generate agent scripts is determining when an agent should start performing their action. [Porteous et al. 2010] used decomposition planning to determine approximately when and where an actor should be. However, by using a representation such as PAR [Bindiganavale et al. 2000] we can create a loose ordering, where agents know to perform actions based on the previous agents' actions. When describing a scene, it is common to use temporal phrasing such as *Then this happened*. We use this information for turn taking. In the absence of temporal phrasing, we assume multiple agents perform an action concurrently with the last action encountered in the script.

Actions within a story can also be in progress when the scene begins. For example, *Anna is holding the cup* implies that the character *Anna* is performing a *holding* animation when the user enters into the story. Other actions more obviously occur during a scene, such as *George runs to the bathroom*. To distinguish these two for rendering purposes, we examine the tense of the verb. In the two aforementioned sentences, the first is a continuous tense and the

second is simple present. Continuous actions are realized in the same manner as active verbs, but simply start their realization before the user is introduced to the scene. If the pre-processing time for the action is known, or it is a pose or event action, then a relative time frame can be established. For gestural actions, without that information, a loop action should be constructed, so that a user may be introduced to the scene at any time and still examine the action.

It is interesting to note from this work that objects that are part of an action frame are not treated any differently than objects that are used to describe a scene. When designing rich scenes with several actions being performed, the placement of objects used by agents is critical. Using our generated agent scripts, it should be possible to control the placement of the objects being interacted with. Future work will determine the ability and usefulness of such a constraint. Another interesting note, also mentioned in [Coyne et al. 2010], is that users may input broad categorizations of spaces and expect the system to furnish those appropriately. While rules can be easily written for some rooms such as a kitchen or bedroom, other rooms may be contextual based on the object models available. Automatically populating these rooms with objects that do not matter to the story provide visual meaning and help set the scene. In the future we plan to create a corpus of room layout data in order to statistically populate scenes with a set of default objects. These objects may be used directly in the story, and will provide a more meaningful and believable environment.

## References

- BINDIGANAVALA, R., SCHULER, W., ALLBECK, J. M., BADLER, N. I., JOSHI, A. K., AND PALMER, M. 2000. Dynamically altering agent behaviors using natural language instructions. In *Proceedings of the fourth international conference on Autonomous agents*, ACM, New York, NY, USA, AGENTS '00, 293–300.
- COYNE, B., AND SPROAT, R. 2001. Wordseye: An automatic text-to-scene conversion system. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, SIGGRAPH '01, 487–496.
- COYNE, B., RAMBOW, O., HIRSCHBERG, J., AND SPROAT, R. 2010. Frame semantics in text-to-scene generation. In *Knowledge-Based and Intelligent Information and Engineering Systems*, Springer Berlin Heidelberg, R. Setchi, I. Jordanov, R. Howlett, and L. Jain, Eds., vol. 6279 of *Lecture Notes in Computer Science*, 375–384.
- FÖRGER, K., HONKELA, T., AND TAKALA, T. 2013. Impact of varying vocabularies on controlling motion of a virtual actor. In *Intelligent Virtual Agents*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds., vol. 8108 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 239–248.
- MA, M. 2006. *Automatic Conversion of Natural Language to 3D Animation*. PhD thesis, University of Ulster.
- PORTEOUS, J., CAVAZZA, M., AND CHARLES, F. 2010. Applying planning to interactive storytelling: Narrative control using state constraints. *ACM Trans. Intell. Syst. Technol.* 1, 2 (Dec.), 10:1–10:21.

\*e-mail: jbalint2, ygingold, jallbeck@gmu.edu