

# Building 3D Virtual Environments from Natural Language Descriptions: Text-to-Scene Conversion

KEVIN GLASS AND SHAUN BANGAY

---

## 1. INTRODUCTION AND RELATED WORK

The conversion of fictional books into digital format presents many opportunities for the computational study of natural language, particularly in fictional literature. One such opportunity is in the field of virtual reality. Our research aims to determine to what extent the book can be used to provide information regarding the layout of a three dimensional scene. It is our hypothesis that electronic fictional books provide an opportunity to create rich virtual environments with greater ease than current manual systems allow.

Related projects in Text-to-Scene conversion include the Put System [Clay and Wilhelms 1996], WordsEye [Coyne and Sproat 2001], and the Story Picturing Engine [Joshi et al. 2004]. Text-to-Scene conversion also draws knowledge from a number of fields, including 3D Computer Graphics, Computational Linguistics, Cognitive Psychology and Knowledge Representation.

## 2. PROCESS DESIGN

Several aspects of Text-to-Scene conversion have been identified which can benefit from automation:

- Script Extraction*: correctly extracting direct-speech from the text, and associating it with the correct speaker. The desired result is a script which can be easily understood by a speech synthesis system.
- Scene Specification*: the extraction of useful information regarding the contents, layout and attributes of objects/characters in the scene.
- Sequencing*: constructing a coherent order of events that can be reconstructed in a virtual environment.

The Text-to-Scene conversion process is designed as follows: A digital book, sourced from the Gutenberg Corpus is passed through a text-tokenisation process, and then through a number of Parts-of-Speech taggers [Glass and Bangay 2005], which identify the word-class of each word. The text is also passed through a syntactic parser in order to determine meta-data such as *subject* and *object* of a sentence. The meta-data derived from the taggers and parsers is then consolidated into a single taxonomy, which is the core data source for constructing 3D scenes.

## 3. COMMENTS AND FURTHER WORK

The consolidation of data derived from taggers and parsers proved to be a challenging task. However, the incorporation of the data into a single taxonomy has been achieved. Experiments are currently being performed to identify the best method for Script Extraction. These tests include the use of WordNet [Fellbaum 1998] to provide semantic relations between words. In addition, a standardised, annotated 3D object library is also under construction which provides 3D models for placement in a virtual world.

## REFERENCES

- CLAY, S. R. AND WILHELMS, J. 1996. Put: Language-based interactive manipulation of objects. *IEEE Computer Graphics and Applications*, 31–39.
- 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 Press, 487–496.
- FELLBAUM, C., Ed. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- GLASS, K. AND BANGAY, S. 2005. Evaluating parts-of-speech taggers for use in a text-to-scene conversion system. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*. White River, South Africa.
- JOSHI, D., WANG, J. Z., AND LI, J. 2004. The story picturing engine: finding elite images to illustrate a story using mutual reinforcement. In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*. ACM Press, New York, NY, USA, 119–126.