**Computational Stylistics: Generation, Classification and Transformation**

**Foad Khosmood, Robert Levinson**

Department of Computer Science, Jack Baskin School of Engineering, UC Santa Cruz

---

### Overview

Natural Language style have existed for a long time in both spoken and written form. There is not real consensus as to what precisely defines a language style or to what extent style is separable from semantics of the text or speech. We find definitions from Jane Wodleb (1969) and Jesi Kalpyra (2004) world.

A style is a consistent and distinguishable tendency to make some [of these] linguistic choices. Style is on the surface level, very obviously distinguishable as the choice between items in a vocabulary, between types of textual construction, between the various ways a text can be written from the material it is made of (Krutick, 2004).

Stylistic analysis has been done by scholars in literary studies and humanities. Computational stylistics has been used for digital forensics, authorship attribution, natural language generation (NLG) and interactive entertainment. Our present work considers three specific questions in stylistics:

1. How can we reliably generate language in a chosen style?
2. How can we recognize a style of language or classify texts according to known styles?
3. How can we transform language from one known style to another without losing meaning or coherence?

### Style-based Classification

#### Style Markers

- Measurable and deterministic textual/linguistic features of a corpus
- Antigonic to the size of the corpus and number of documents
- Many researchers have experimented with many different markers
- We propose a 5 level style taxonomy based on marker types used in literature
- Taxonomic proposal currently under peer review (2011)

#### Proposed Style Taxonomic Hierarchy

- Concept
- Family
- Marker
- Parameter
- Quantifier

### AAAC-04 Corpus

- Material from Ad-hoc Authorship Attribution Contest in 2004
- 13 Classification problems (A-H), each with a set of labeled documents associated with an author and set of unlabeled documents
- Task is to match up the unlabeled documents with authors
- Java Graphical Authorship Attribution Program (JGAAAP) created and supplied with AAAC documents, as well as event sets and classification algorithms from the contest
- Created by professor Patrick Juurl, Duquesne University
- We extracted 136 Marker-quantifiers from the AAAC corpus documents for English language problems A-H, representing about 90 document classification problems

#### Relevant Publications

- Kalgren, Agnostic to the size of the corpus and number of documents
- Kalgren, Stylistic Natural Language Generation
- Kalgren, The Grapevine System
- Kalgren, The Grapevine System
- Kalgren, The Grapevine System
- Kalgren, The Grapevine System

---

**Stylistic Natural Language Generation**

- Generates a series of character-gossip dialogues
- System accepts personality profiles of characters and belief statements about the world
- State-machine driven gossip-conversation speech acts
- Two levels of speech variations:
  1. State-machine branching and dialogue plan decisions based on personality
  2. Sentence planning and stylistic enrichment using PERSONAGE (Meronese and Walker)

**The Grapevine System**

- Generates a series of two-character gossip dialogues
- System accepts personality profiles of characters and belief statements about the world
- State-machine driven gossip-conversation speech acts
- Two levels of speech variations:
  1. State-machine branching and dialogue plan decisions based on personality
  2. Sentence planning and stylistic enrichment using PERSONAGE (Meronese and Walker)
- Validation using user evaluation surveys (50 users)
- Results published in:

---

**Style Transformation**

### Transforms

- Are ways to re-write, paraphrase or alter the text, implemented as functions
- Most perform simple sentence level functions without consideration for larger style of the document
- Stylistic impact measure by shift in classification

### Relevance

- The neural network replaces the transforms in an event set
- The transform ran on 13 documents in AAAC-04 problem A
- Using standard JGAAAP settings and FNCC algorithm

---

**Relevant Publications**