



Poetry Generation

Introduction to Computational Creativity

Jukka Toivanen



Poetry

- Essentially a very sophisticated way to convey thoughts and feelings with tight interaction between content and form
- Usually short (compared to stories, novels)
- More freedom regarding grammar
- But also more requirements for the form (and content)
- Many different criteria for the aesthetic value
- One of the most studied forms of creative NLG



Research Goals?

- Test bed for NLP, data mining and machine learning methods
- Ultimate AI challenge
- Artistic results on their own
- Some day also better understanding of creativity and human thought processes (cognitive modeling)



Different Ways to Handle Syntax

- Fixed templates
- Generative grammars
- Statistical language modeling from existing texts
- Assembly from text fragments
- ...



Different Ways to Handle Semantics

- None (just statistical language modeling)
- Association-based analysis and lexical selection
- Assembly from fragments with known semantics
- Formal logical semantics



Existing Systems

Several systems have been developed:

- ASPERA (Gervás 2001): Case-based reasoning approach
- The work of Manurung et al (2000): Rich linguistic knowledge based on strong formalisms for syntax, semantics, and phonetics
- Full-FACE poetry generation by Colton et al. (2012)
- Work of Toivanen et al. (2012,2013,2014): Interaction of two distinct corpora for content and form. The approach requires minimal amount of explicit grammar specification
- Many others



Evolutionary Poetry

- Evolutionary algorithms in poetry generation
- McGonagall system by Manurung et al.
 - Grammaticality
 - Meaningfulness
 - Poeticness
- Poetry generation as a state space search problem
- Genetic algorithms to solve this problem



Example

the cat is the cat which is dead;

the bread which is gone is the bread; the cat which consumed

the bread is the cat

which gobbled the bread which is gone



Our Research on Poetry Generation

- Especially, what kind of automatic and maximally unsupervised methods could be used to generate poetry
 - Statistics of word co-occurrences
 - Morphological analysis and synthesis
 - Constraint satisfaction methods
 - Document-specific content



Document-Specific Poetry

- Generating poetry that reflects loosely a specific news story or other document
- Evoking fresh mental images and viewpoints that are related to the document but not exactly contained in it
 - Using document-specific words in poetry generation
 - Extension of the basic word substitution based poetry generation method by Toivanen et al. (2012)



Our Aims in Poetry Generation

- Maximally unsupervised
 - Minimum amount of hand-crafted linguistic, world, and poetry domain knowledge
 - No explicit grammars
 - No manually generated templates
 - No knowledge bases
 - Use of statistical corpus-based methods
- Fresh mental images evoked by novel associations



Benefits and Restrictions of this Approach

- Pros:
 - Flexibility
 - Language independence
 - Direct learning from corpora, minimal amount of hand-crafted rules
- Cons:
 - The quality of the results varies
 - There are still many aspects in the generation procedures that are difficult to control



Word Association Analysis

- Finding content words for replacement poetry
- General association calculation method proposed by Gross et al. (2012)
- Recent extension to document specific associations (Gross, Doucet, and Toivonen 2014)
 - Which word pairs are novel in a specific document?
 - A background corpus as a reference of novelty
 - Contrasting a specific document (called foreground) to a set of documents in the background corpus
 - log-likelihood ratio (LLR) used to measure document-specific word associations



Example News Story

Justin Bieber on Miami drink-drive charge after 'road racing'

Pop star Justin Bieber has appeared before a Miami court accused of driving under the influence of alcohol, marijuana and prescription drugs. Police said the Canadian was arrested early on Thursday after racing his sports car on a Miami Beach street. They said he did not co-operate when pulled over and also charged him with resisting arrest without violence and having an expired driving licence. (...)

BBC News, 23 January 2014



Document-Specific Word Associations

- Descriptive associations could be:
 - “bieber” and “alcohol”
 - “bieber” and “prescription”
 - “justin” and “alcohol”
 - ...
- Not so descriptive associations include:
 - “justin” and “bieber”
 - “sports” and “car”
 - “driving” and “licence”



How to Find Document-Specific Word Associations

Counts in the News Story

	Bieber	\neg Bieber
Alcohol	2	0
\neg Alcohol	4	22

Counts in the Background Corpus

	Bieber	\neg Bieber
Alcohol	0.	19419.
\neg Alcohol	244.	33967685.



How to Find Document-Specific Word Associations

Counts in the News Story

	Justin	\neg Justin
Bieber	3	3
\neg Bieber	0	22

Counts in the Background Corpus

	Justin	\neg Justin
Bieber	5.	239.
\neg Bieber	3747.	33983357.



How to Find Document-Specific Word Associations

Foreground Counts

	w_1	$\neg w_1$
w_2	k_{11}	k_{12}
$\neg w_2$	k_{21}	k_{22}

Background Counts

	w_1	$\neg w_1$
w_2	k'_{11}	k'_{12}
$\neg w_2$	k'_{21}	k'_{22}

$$D_{LLR} = 2 \sum_{i=1}^2 \sum_{j=1}^2 k_{ij} (\log(p_{ij}) - \log(q_{ij})).$$



Document-Specific Word Associations

- Find word pairs whose co-occurrence distribution in the document deviates most from the background corpus
- These words are descriptive for the novel content of the document in question
- Use these words as replacements in the poetry generation phase



Example Associations

Most novel pairs	Least novel pairs
say, beiber	los, angeles
say, police	later, jail
miami, beiber	sport, car
miami, say	car, early
beiber, police	thursday, early
beach, beiber	marijuana, alcohol
beach, police	prescription, alcohol
car, say	sport, thursday
beiber, alcohol	car, street
beiber, los	prescription, marijuana



Overview of the Generation

- Word substitution method as described by Toivanen et al. (2012)
 - A piece of text from a corpus (e.g. poetry)
 - Replacing most of the words with words relevant to a specific news story
 - Morphological analysis and synthesis
 - Stanford POS-tagger (Toutanova et al., 2003)
 - morpha & morphg tools (Minnen, Carroll, and Pearce, 2001)
 - For other languages (Finnish, German, French) other tools needed



Overview of the Generation Process

*Is it the dirt, the squalor,
the wear of human bodies,
and the dead faces of our neighbours?
These are but symbols.*

Project Gutenberg, Imagist Poetry

*Is it the entourage, the sport,
the singer of later lamborghinis,
and the early thursdays of our singers?
These are but justins.*



Experiment

- The corpus from which templates were taken contained mostly Imagist poetry from the Project Gutenberg
- Background corpus was the English Wikipedia
- Several different news stories, e.g.
 - Justin Bieber drinking and driving
 - Huawei profits surging
 - Ukrainian prime minister resigning
 - US states reconsidering execution methods
- The following poems were selected randomly and presented as they are



Huawei Profits Surge...

Oaks

*and impact technologies,
rise buying with transfer, rise:
their comfortable technology.*



Ukrainian Prime Minister Resigning...

And always concrete! Oh, if I could ride

With my week resigned concrete against the repeal

Do you resign I'd have a parliament like you at my television

With your azarov and your week that you resign me? O ukrainian week,

How I resign you for your parliamentary legislation!



US States Reconsidering Execution Methods

I die;

perhaps I have begun; this is a doubt;

this is a prisoner;

and there is state....



Some Examples of Societal Impact

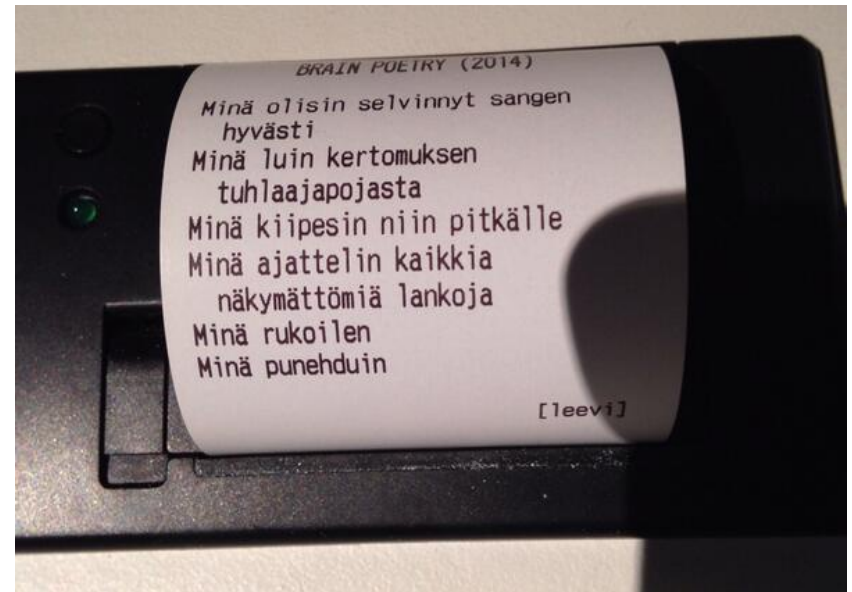


Book with Heikki Paakkanen





Brain Poetry at the Frankfurt Book Fair





Closer Look at the Intentionality

- What to express in a poem?
 - Converting logical formulas into poetry
 - Converting knowledge base information into poetry
 - Conveying semantic relationships in different ways
- Does poetry always aim to express specific information?
 - Which one is more important, content or form?
 - Different literary theories have different views of this (e.g. Russian Formalism vs. Structuralism)



References

- Oskar Gross, Antoine Doucet, Hannu Toivonen. Document summarization based on word associations. In Proceedings of the 37th international ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2014.
- Anna Kantosalo, Jukka M. Toivanen, Ping Xiao, Hannu Toivonen: From Isolation to Involvement: Adapting Machine Creativity Software to Support Human-Computer Co-Creation. Proceedings of the Fifth International Conference on Computational Creativity, 2014.
- Jukka M. Toivanen, Hannu Toivonen, Alessandro Valitutti, Oskar Gross: Corpus-Based Generation of Content and Form in Poetry. Proceedings of the Third International Conference on Computational Creativity, 175-179, 2012.
- Jukka M. Toivanen, Matti Järvisalo, Hannu Toivonen: Harnessing Constraint-Programming for Poetry Composition. Proceedings of the Fourth International Conference on Computational Creativity, 160-167, 2013.
- Jukka M. Toivanen, Oskar Gross, Hannu Toivonen: "The Officer is Taller Than You, Who Race Yourself!" Using Document Specific Word Associations in Poetry Generation. Proceedings of the Fifth International Conference on Computational Creativity, 2014.