Automatic, literary text generation by artistic indexes

Mohammad Rahmani

1 Introduction

While numerous creative language generation studies, have focused on topics such as generating poems, stories, etc by learning from the corpora, few have addressed incorporation of artistic indexes (such as antonym shift, ambiguity and VAD gravity as described in Section 3.2) in their generation process. The few which have done so, have counted on training the model to learn artistic styles from the corpora which results in repetitive results similar to the structures in the corpora. As such, a model that can generate literary text by considering artistic indexes is missing. Particularly ambiguity and antonym shift which base most of other artistic merits have never been used in an automated text generation process. This proposal first suggests application of semantic vector spaces to mathematically defining some artistic indexes and then application of stochastic gradient descent to generate language which incorporate maximum of the aforementioned artistic indexes. Regarding the terminology used in this proposal, it is worth mentioning that art is differentiated from creativity in the sense that some artistic indexes such as nostalgia which is based on a prior emotional experience may contradicts creativity which is the matter of introducing novelty and unexpectedness in a work of art (Jordanous, 2012). Aesthetics which is the matter of proportionality of the constituents with a whole, can be considered as only a potential constituent of a work of art and not the base of it. As such, none of these two terms can substitute the term "artistic merit" in this proposal. As such, this proposal[1] tries to address the absence of a creative, artistic text generation model by selecting dependency sequences derived from corpora such that the maximum artistic merit is achieved through mathematically defining a set of artistic indexes using stochastic gradient descent.

[1]A piece of code in Java (just for the sake of clean code Java is chosen to build the skeleton, otherwise Python or a mix of two could also be used) which is being updated continuously is available at this git repository to help with better understanding the objective and the methodology:
`https://github.com/donkarlo/gissoo`.
Additionally a UML class diagram is available at
`https://github.com/donkarlo/gissoo/blob/master/docs/assets/artisticness.png`

2 Objective

If we have $E = pond, frog$ and $p = tobe, jump$, then this proposal plans to generate a work of art by randomly constituting dependencies such as "pond is silent", "frog jumps in the pond", "pond is not silent"[2] such that maximum artistic merit can be achieved.

Taking D in the following equation as a set of extracted potential dependencies in form of subjects, predicates and objects from corpora

$$\begin{aligned} if\ E = \{e_1, ..., e_q\}\ and\ P = \{p_1, ..., p_q\}\ then \\ D = \{d_f = (s_i, v_h, o_j) | s_i, o_j \in E, v_h \in V\} \end{aligned} \tag{1}$$

where E is the set of extracted entities(subjects and objects) and V is the set of extracted predicates, then this proposal is trying to introduce dependency sequences such as:

$$l_n = (d_t)_{t=1}^{n} \tag{2}$$

such that $a(l_n)$ in

$$a(l_n) = \sum_{k=1}^{m} a_k(l_n) \tag{3}$$

maximizes where a_k is a single artistic merit index member in A as follows:

$$A = \{a_k : l_n \longrightarrow (0, 1) | k \in \{1, ..., m\}\} \tag{4}$$

In Section 4.4, a solution based on stochastic gradient will be suggested to study in this proposal to generate diverse sequences of dependencies such as the one presented in Equation 2 with local maximum artistic merit values in each run.

3 Related works

The following subsections are subjects which must be addressed in Section 4 as a part of the solution to accomplish the objectives of this proposal. This section is particularly important since not only it references previous works which might be used in Section 4, but also, incorporates the definition of some terminologies applied in the rest of this proposal.

3.1 Creative language generation

Literary text generation, particularly, poetry has been the subject of many studies Oliveira (2017); Lamb et al. (2017). The closest literary form to the expected

[2]Inspired from a haiku by Yosa Buson
An old silent pond...
A frog jumps into the pond,
splash! Silence again.

results of this research proposal are poems (Particularly haiku), but it does not put emphasis on any particular literary form. Most poem generation models learn and repeat artistic patterns from corpora such as Daza et al. (2016). Some researchers tried to develop models to avoid such repetitions in favor of creativity (Wu et al., 2019). Narratives in form of short stories are also close to such theme, specially the idea of extracting possible dependencies between objects in corpora to form stories from a subset of this, is presented in works such as McIntyre and Lapata (2009) which introduces a rank-based solution to select dependency sequences with arbitrary lengths and psycho-linguistic indexes to evaluate the results. Moreover McIntyre and Lapata (2010) uses evolutionary algorithms instead of the rank-based approach introduced in McIntyre and Lapata (2009). Additionally, interesting researches have been carried out by Veale (2013b,a); Gervás (2009); y Pérez R (2015).

3.1.1 Dependency extraction

Although Standford Dependencies (Schuster and Manning, 2016) will be used to extract dependencies from corpora, yet other innovative solutions have been proposed. An approach to extract chains of dependencies as constituents of a story is proposed in (Chambers and Jurafsky, 2008). Martin et al. (2018) suggests an interesting neural approach for the same type of extraction for story generation. de Marneffe et al. (2014) suggests an improved taxonomy to capture grammatical relations across languages, including morphologically rich ones. Nivre et al. (2016) has discussed solutions to extract dependencies further than subject-object relations.

3.2 Artistic indexes

In this section three artistic indexes are introduced which will be addressed in Sections 4.2.1, 4.2.2 and 4.2.3 as sub-sections of the suggested methodology. Certainly more indexes could be investigated, yet for the sake of feasibility of accomplishing the objectives of this proposal within the reasonable time for a PhD, just the following three artistic indexes will be studied.

3.2.1 Transition between antonym semantic vector classes

Various transition between different or even antonym semantic classes is considered as a source of artistic merit in several literary forms. Haiku is one such form (Figure 1) Ono et al. (2015) suggests a solution for antonym detection by word embedding.

3.2.2 Ambiguity

Ambiguity is an appealing constituent of an work of art (Muth et al., 2015). For example, it is not determinable whether Mona Lisa is smiling and if she is, is it a bitter or a sweet smile? If the course of the events (dependencies) take the final perceived semantic to the border of two different semantic classes,

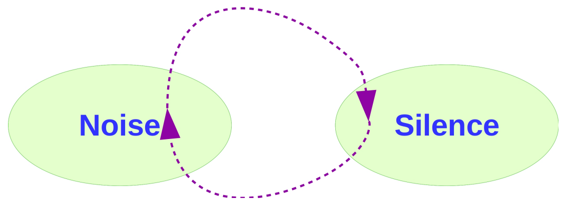

Figure 1: Shift between antonym semantic classes. For example: [The] pond is silent. [A] Frog splash into it. [The] pond is silent [again]

then such ambiguity provokes imagination (Figure 2). A subsequent result of ambiguity is suspense. Cheong (2006) suggests a model to generate suspense. O'Neill and Riedl (2011) presents a suspense detection system based on the correlation between perceived likelihood of a protagonist's failure and the amount of suspense reported by the audience.

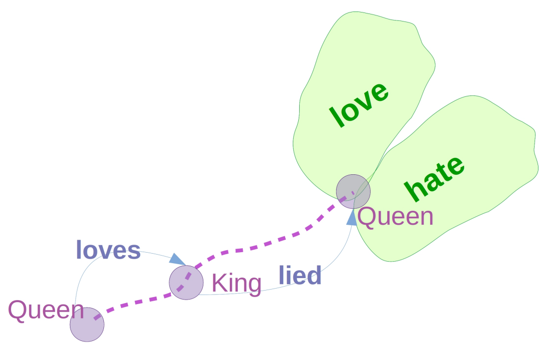

Figure 2: A dependency sequence which is ended in an ambiguity between "love" and "hate" in a story such as: "The queen loves the king, but the king lied the queen."

3.2.3 Valence, arousal and dominance (VAD) gravity of dependency sequences

Whenever the course of the events takes the situation to a semantic class from which escaping is impossible then the audience mind deeply engages with imagining potential break outs. For example, if the course of the events in the story, ends in a deep regret (Figure 3), then however much a story character commits possible predicates, s/he can not change the semantic class from regret to something different (such as happiness ....). Naturally, such endings engages the imagination of the audience (reader) to think of ways with which the character can release himself from such situations. Buechel et al. (2016) suggests a method to assign a VAD value to a sequence of sentences. Studies such as Agrawal et al. (2018); Mao et al. (2019) and Li et al. (2017) have regarded VAD from a semantic vector perspective. Li et al. (2017) suggests an embedding particular to affective meaning which includes the arousal and valence of the affection.

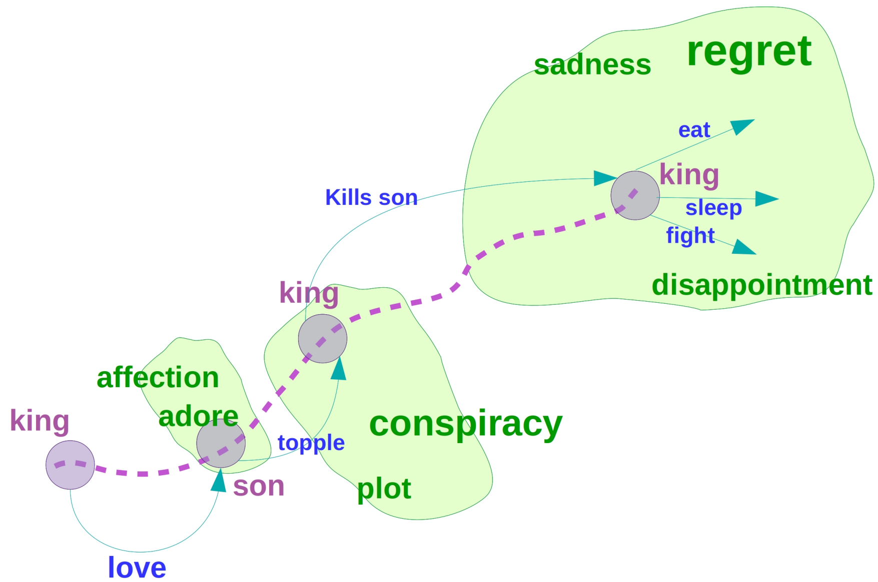

Figure 3: Dependency sequence ultimately shifts to a semantic class(regret) in a story such as "The king loves his son. The son wants to topple the king. The king kills his son." with such a huge VAD that no other dependency in D can pull it out.

3.3 Dependency (event) embedding

There are some solutions which have been proposed toward event embedding such as Weber et al. (2018). Taking a dependency in D as an event, such solutions are also applicable for dependencies.

3.4 Application of stochastic gradient for text generation

Previous to this proposal, it seems that no other study has used stochastic gradient descent to generate a dependency sequences with maximum artistic merits. Previously, studies such as (Welleck et al., 2019b,a,b) have used neural approaches in generation of diverse texts with similar inputs, yet stochastic gradient descent, due to its nature seems to be also a feasible solution for such objectives.

3.5 Surface realization

Many surface realizers have been developed in recent years (Gatt and Reiter, 2009; Belz et al., 2012) but simple old solutions such as Lavoie and Rambow (1997) can be used for primary tasks in this proposal.

4 Methodology

Briefly, the methodology is composed of choosing a textual corpora from which a corpus of dependencies could be extracted and then introducing dependency sequences from D such that they locally maximize the artistic indexes introduced in Section 3.2 by taking advantage of stochastic gradient descent. In details:

4.1 Corpora building

Choosing textual Corpora A corpora such as Wikipedia which doesn't inherently incorporate literary text will be used for dependency extraction. The reason for such a choice is to remove the bias that the model is learning its artistic merit values from the corpora.

Dependency parsing Stanford coreNLP dependency parser will be used to extract dependencies from corpora to form D in Section 2. If necessary, other neural models (such as those introduced in Section 3.3) will also be used for more complex requirements.

4.2 artistic indexes

In this phase a set of artistic indexes will be mathematically defined so that sequences such as l_n in Equation 2 can be mapped to corresponding artistic merit values between 0 and 1 as described in Equation 3.3.

4.2.1 Transition between antonym vector semantic classes

First, a semantic vector space will be developed such that antonym words or dependencies are placed in different classes (As a starting point, previous studies such as Ono et al. (2015) will be considered). Then, a model must be developed or trained to detect the shifts between antonym semantic classes in a sequence of dependencies.

4.2.2 Ambiguity

First a clustering model should be developed to cluster the words such that on the one hand, predicates can shift perceived meanings by the reader from dependencies in a sequence to a vector-defined position in a final residential class (such as Figure 2) and on the other hand, words with similar meanings cluster into same classes. Another model should be developed to measure the distance between the landing site of a dependency sequence meaning in its residence class and its neighboring classes. A dependency sequence with average closer distance to more classes is considered more ambiguous and consequently more artistic.

4.2.3 Valence, arousal and dominance (VAD) gravity of dependency sequences

A semantic vector space should be developed such that the resulting classes can adjust their extension according to the course of dependency sequences so that the size of predicate vectors comply in a way that irrelevant predicates to the situation lose size while others grow in size.

4.3 Determining a model for $a(l_n)$

This could be simply done by assigning the summation of occurrences of each artistic index in l_n or could be any machine learning model such as a neural network that gets the rate of the antonym shifts or ambiguities in l_n as input and outputs a number in $(0, 1)$ as output. The training data for the former approaches could be gained from crowd sourcing techniques such as Amazon Mechanical Turk. Also there is a possibility that different values of correlation exists between different artistic indexes. For example, studies can be carried out on the degree by which two classes of a word embedding space are antonyms and the degree by which ambiguity is introduced according to the distance of the final landing location of predicates to vector space class border and to see if these two negatively or positively correlated.

4.4 Stochastic gradient descent for dependency sequence generation

A dependency(event) embedding such as

$$\begin{aligned} b : d_t &\longrightarrow \mathbb{R}^z \\ d_t &\longrightarrow \vec{d_t} \end{aligned} \tag{5}$$

will be developed and will be used (using one of the solutions mentioned in Section 3.3) to map dependencies in D to vectors. Taking advantage of stochastic gradient descent and the three aforementioned artistic indexes, a random or arbitrary dependency vector such as $\vec{d_i} \in \vec{D}$ will be given to a stochastic gradient descent function for maximizing Equation 3. The next dependency embedding will be chosen by stochastic gradient descent along the partial derivative direction until $a(l_n)$ reaches a local maximum. As such, a sequence of dependencies will be introduced with maximum artistic indexes (Figure 4). Since the first dependency is chosen arbitrarily or randomly, then every time the first dependency changes, a new trajectory(sequence) of dependencies toward the local artistic merit maximum is introduced. Such trajectory can be even more diversified when Equation 3 has more than one local maximum. As stochastic gradient descent's steps number toward reaching a maximum values is not predictable, then another advantage of this approach is generation of dependencies with unknown lengths. It is noticeable that the three artistic indexes may contradict each other. In other words, in some cases they defuse each other. For example, ambiguity appeal may inherently defuse antonym shifts because one is based on ending a dependency sequence between two contradicting semantic classes while the other is the matter of landing dependency members in distinct antonym classes. As such, existence of local maximums seems to be inevitable.

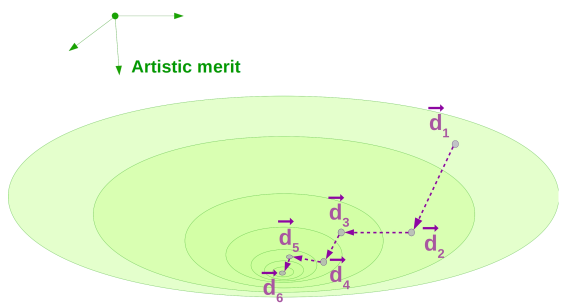

Figure 4: Dependencies from $\vec{d_1}$ to $\vec{d_6}$ where the maximum artistic merit of all indexes in Section 4.2 happen. If $\vec{d_1}$ was happen to be chosen in another location or the stochastic gradient descent on Equation 3 repeats, certainly another sequence of dependencies with a different length would be introduced.

4.5 Surface realization from dependencies

In this phase, the dependency sequences generated in Section 4.4 will be converted to real natural language sentences by either taking advantage of methods introduced in Section 3.5 or more approaches tailored to the particular requirements of this proposal.

5 Baselines and evaluation

5.1 Baseline

McIntyre and Lapata (2009, 2010) introduce psycho-linguistic methods as an index for interestingness. Their solution fits as a baseline model for this proposal because it also functions based on joining dependencies derived from corpora by either a rank or evolutionary based models to choose the appropriate dependent sequence among a set of dependency population such as D.

5.2 Artistic merit index detector models

If the model in Section 4.4 generates n dependency sequences then:

Antonym class shifts detector model: If the antonym detector in the model in Section 4.2.1 detects t haiku with antonyms, (considering the fact that all haiku include antonym semantic class shift) then we expect

$$n \times \frac{|A|}{t} \tag{6}$$

of dependency sequences generated by the method in Section 4.4 must include at least one antonym shift other wise the model must be improved.

Ambiguity and VAD gravity The same rules or antonym detector model, applies to ambiguity and VAD gravity models, although finding corpora for these two indexes is more challenging.

5.3 Evaluation

Human evaluation Human evaluation is not simply replaceable. Here, machine evaluation techniques which were described in Section 5.2 will be also used since artistic merit qualities are hard to judge by human agents according to their understanding of art. But since the proposed approach is based on mathematical definition of artistic indexes which are more strict but human evaluation will be used extensively to find out how close machine and human judgments are in this area.

Surface realization Using methods such as BLEU (Papineni et al., 2002) or METEOR (lav) the final results in natural language could be evaluated. Yet, human evaluation on final results is also an asset for overall assessment of the model's performance.

6 Timetable

- Phase 1: 5 months: literature review.
- Phase 2: 4 months: Choosing the corpora and dependency extraction from an unbiased corpus
- Phase 3: 8 months: Developing models to cluster words antonym words into different semantic classes and detect semantic antonym shifts in a sequence of dependencies
- Phase 4: 9 months: Developing models that can cluster words according to their meaning and detect ambiguity appeal of a sequence of dependencies

- Phase 5: 8 months: Developing a model that can detect historical VAD of the final semantic class in which a sequence of dependencies has resided.
- Phase 6: 8 months: Stochastic Gradient descent implementation to generate dependency sequences with maximum artistic indexes described in three previous phases.
- Phase 7: 6 months: making improvements iterations in previous stages and trying to add supplementary artistic indexes.

7 Papers

During the PhD these papers should be published to accomplish the objective and the road map sketched in methodology section.

- A paper will be developed to present the results of clustering antonym words into different classes and detecting semantic antonym shifts at the end of phase two. Probably a collection of haiku will be used to determine the performance of the antonym shifts detector model and present the results.
- A paper will be developed to present the results of clustering the word embedding according to their meanings and detecting the ambiguity appeal of dependency sequences.
- A paper will be developed to present the result of the VAD semantic vector space.
- A paper with reference to three previous papers will be developed to present the results of dependency sequences introduced by stochastic gradient descent, introduced in Section 4.4 and evaluated according to the metrics described in Section 5.

References

METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments.

Ameeta Agrawal, Aijun An, and Manos Papagelis. Learning emotion-enriched word representations. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 950–961, 2018.

Anja Belz, Bernd Bohnet, Simon Mille, Leo Wanner, and Michael White. The surface realisation task: Recent developments and future plans. In *INLG 2012 - Proceedings of the Seventh International Natural Language Generation Conference, 30 May 2012 - 1 June 2012, Starved Rock State Park, Utica, IL, USA*, pages 136–140, 2012.

Sven Buechel, Johannes Hellrich, and Udo Hahn. Feelings from the past - adapting affective lexicons for historical emotion analysis. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities, LT4DH@COLING, Osaka, Japan, December 2016*, pages 54–61, 2016.

Nathanael Chambers and Daniel Jurafsky. Unsupervised learning of narrative event chains. In *ACL 2008, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, USA*, pages 789–797, 2008.

Yun-Gyung Cheong. A computational model of narrative generation for suspense. In *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*, pages 1906–1907, 2006.

Angel Daza, Hiram Calvo, and Jesús Figueroa-Nazuno. Automatic text generation by learning from literary structures. In *Proceedings of the Fifth Workshop on Computational Linguistics for Literature, CLfL@NAACL-HLT 2016, June 16, 2016, San Diego, California, USA*, pages 9–19, 2016.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. Universal stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.*, pages 4585–4592, 2014.

Albert Gatt and Ehud Reiter. Simplenlg: A realisation engine for practical applications. *Proceedings of the 12th European Workshop on Natural Language Generation, ENLG 2009*, pages 90–93, 04 2009.

Pablo Gervás. Computational approaches to storytelling and creativity. *AI Magazine*, 30(3):49–62, 2009.

Anna Jordanous. A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3):246–279, 2012.

Carolyn Lamb, Daniel G. Brown, and Charles L.A. Clarke. A taxonomy of generative poetry techniques. *Journal of Mathematics and the Arts*, 11(3): 159–179, 2017.

Benoit Lavoie and Owen Rambow. A fast and portable realizer for text generation systems. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, ANLC '97, pages 265–268, 1997.

Minglei Li, Qin Lu, Yunfei Long, and Lin Gui. Inferring affective meanings of words from word embedding. *IEEE Trans. Affective Computing*, 8(4):443–456, 2017.

Xingliang Mao, Shuai Chang, and Jinjing Shi. Sentiment-aware word embedding for emotion classification. 2019.

Lara J. Martin, Prithviraj Ammanabrolu, Xinyu Wang, William Hancock, Shruti Singh, Brent Harrison, and Mark O. Riedl. Event representations for automated story generation with deep neural nets. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 868–875, 2018.

Neil Duncan McIntyre and Mirella Lapata. Learning to tell tales: A data-driven approach to story generation. In *ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2-7 August 2009, Singapore*, pages 217–225, 2009.

Neil Duncan McIntyre and Mirella Lapata. Plot induction and evolutionary search for story generation. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*, pages 1562–1572, 2010. URL `http://www.aclweb.org/anthology/P10-1158`.

Claudia Muth, Vera M. Hesslinger, and Claus-Christian Carbon. The appeal of challenge in the perception of art: How ambiguity, solvability of ambiguity, and the opportunity for insight affect appreciation. 2015.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan T. McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016.*, 2016.

Hugo Gonçalo Oliveira. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *INLG*, 2017.

Brian O'Neill and Mark Riedl. Toward a computational framework of suspense and dramatic arc. In *Affective Computing and Intelligent Interaction - 4th International Conference, ACII 2011, Memphis, TN, USA, October 9-12, 2011, Proceedings, Part I*, pages 246–255, 2011.

Masataka Ono, Makoto Miwa, and Yutaka Sasaki. Word embedding-based antonym detection using thesauri and distributional information. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA.*, pages 311–318, 2002.

Sebastian Schuster and Christopher D. Manning. Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*, 2016.

Tony Veale. Creativity as a web service: A vision of human and computer creativity in the web era. In *Creativity and (Early) Cognitive Development: A Perspective from Artificial Creativity, Developmental AI, and Robotics, Papers from the 2013 AAAI Spring Symposium, Palo Alto, California, USA, March 25-27, 2013*, 2013a.

Tony Veale. Once more, with feeling! using creative affective metaphors to express information needs. In *Proceedings of the Fourth International Conference on Computational Creativity, Sidney, Australia, June 12-14, 2013.*, pages 16–23, 2013b.

Noah Weber, Niranjan Balasubramanian, and Nathanael Chambers. Event representations with tensor-based compositions. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4946–4953, 2018.

Sean Welleck, Kianté Brantley, Hal Daumé III, and Kyunghyun Cho. Non-monotonic sequential text generation. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 6716–6726, 2019a.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. *ArXiv*, abs/1908.04319, 2019b.

Chao-Chung Wu, Ruihua Song, Tetsuya Sakai, Wen-Feng Cheng, Xing Xie, and Shou-De Lin. Evaluating image-inspired poetry generation. In *Natural Language Processing and Chinese Computing - 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9-14, 2019, Proceedings, Part I*, pages 539–551, 2019.

Pérez y Pérez R. From mexica to mexica-impro: The evolution of a computer model for plot generation. *T.R. Besold, M. Schorlemmer, A. Smaill (Eds.), Computational Creativity Research: Towards Creative Machines, Atlantis Thinking Machines 7.*, 2015.