Modelling Interestingness: Stories as L-Systems and Magic Squares

Cosimo Palma[1,2]

[1]University of Naples "L'Orientale", Via Duomo, 219, 80139 Napoli, Italy
[2]University of Pisa, Lungarno Antonio Pacinotti, 43, 56126 Pisa, Italy

Abstract
The purpose of this paper is to attract attention towards modelling the story's *interestingness*, an aspect of computational narratology which has been until now mainly linked to the lexical niveau, or even confined as user-centered evaluation metrics. The approach proposed hereby is grounded on the episode (or *narrateme*, in the sense of Propp); furthermore, it aims at objectivity, i.e. independence from user preferences, by drawing from recent findings in psycholinguistics, information theory, quantitative text analysis and cognitive psychology. A formalism used for the representation of recursive (biological or mathematical) shapes is applied to fairy-tales; one of its mathematical developments, the magic square, is suggested to represent, due to its unique properties, an appropriate mathematical object for modelling the distribution of events' self-information values in a fairy-tale.

Keywords
Interestingness measures, Magic squares, Lindenmeyer systems, Computational narratology, Narrative Representation Models

1. Introduction

The introduction of new narrative devices is triggered by the inherent need for novelty characterising the whole literary production, probably associated, like other human activities, with the dopamine D4 receptor gene [1]. The quest for novelty, however, does not need inspiration when concerning Automatic Story Generation, whereby a clear and sound model of interestingness is required instead [1]. The *narratemes* (recurring motives in simple tales, labelling event's function in the development of the narration, such as, for example, the *Transfiguration* or the *Marriage*) underlying a large sample of Russian folktales was first discovered by Vladimir Propp in his seminal work [2]. Because of the regularity of the event sequence ordering, his comparative analysis could be defined as *morphology*, in relationship with other domains, such as botany, where plants of the same species share a similar shape. Despite the effectiveness of this model in matter of fairy-tales, the same would not be easily applied to text formats of higher length, such as the *novel*, whose atomic components cannot be labelled by one single term.

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'23 Workshop, Dublin (Republic of Ireland), 2-April-2023
cosimo.palma@phd.unipi.it (C. Palma)
https://glottocrisio.github.io/bupbengoc/ (C. Palma)
0000-0002-8161-9782 (C. Palma)

[1]Arguably, the modelling of *beauty*, the ideal feature of any artistic product and Holy Grail of Computational Creativity, can be undertaken only after having achieved this preliminary step.

2. Related work

The *surprisal* or *self-information* is a unit of information theory, that aims to measure the probability of an event according to its specific context [1]. The more expectable an event is, the less surprising it becomes[ibid.]. Therefore, the surprisal is inversely proportional to the probability that an event occurs: $wow(i) = log_2 \frac{1}{(p(i)|Kontext)}$.
This dependency of the expected value of an event from the previous model sets the surprisal within the framework of the conditional probability, where the new data observation D carries no surprise if it leaves the observer believes unaffected, i.e. if the posterior is identical to the prior [3]. Conversely, D is surprising if the posterior distribution resulting from observing D significantly differs from the prior distribution. Therefore we can formally measure surprise elicited by data as some distance measure between the posterior and prior distributions, which can be achieved, among others, using the relative entropy or Kullback-Leibler (KL) divergence. Thus, surprise is defined by the average of the log-odd ratio: $KL(P(M|D), P(M)) = \int P(M|D)log\frac{P(M|D)}{P(M)}dM$.
Other statistical test statistics, such as the χ^2, the *Kolmogorov-Smirnov* and the *Epps-Singleton*, conceived to measure how similar two samples are (i.e. how probable it is that the they are generated by the same distribution function) could theoretically be used for the same purpose.
The concept of *interestingness* has been differently defined in accordance to the discipline it occurs in. In Hilderman & Hamilton (1999) [4] a thorough survey of all possible measures is performed in the field of Knowledge Discovery (KD)[2]. One of the most relevant for our domain is the Silbershatz and Tuzhilin's Interestingness, which determines the extent to which a soft belief is changed as a result of encountering new evidence[3].
Hatzel [5] formulates the *relevance* as the amount of change introduced from an event, whereby the event's type describing the state change receives the highest value, followed by other less relevant (i.e., less *eventful*) categories[4]. Strictly concerning Natural Language Processing (NLP) tasks, it is mostly interpreted as a relative measure, such as the degree of similarity among topics occurring in a documents sample and the user interest, calculated upon web browser logs linked to specific users [6, 7]. A spectacular attempt to capture objective interestingness has been performed by Reagan and colleagues [8], where it is encapsulated within the concept of emotion[5]: the emotional reaction of a sample of readers to some famous novels harvested from the Project Gutemberg[6] has been detected and measured by hedonic analysis[7]. The study has been useful to assess that pivotal events arouse very similar reaction among all test takers. This huge degree of inter-subjectivity, the enormous expenses necessary to carry out the experiment, as well as the difficulties related to the dataset construction for the other studies, leads us to

[2]They are ranked by *representation* (the dataset format on which they are to be applied), *foundation* (probabilistic, distance-based, etc.), *scope* (single rule/rule set) and *class* (objective/subjective).
[3]A *soft belief* is one that an agent is can easily change provided that new evidence is encountered[ibid.].
[4]This approach is particularly noteworthy because it does not require a large annotated training dataset upon which calculating expected values.
[5]The emotion, for simplicity's sake, is therein categorised as *positive* or *negative.* Its intensity, not connotation, is strictly correlated to what in this paper is meant by *interestingness.*
[6]https://www.gutenberg.org/ .
[7]The facial expressions of study participants have been monitored during the reading, and assigned to a value.

the question, of weather it be possible to model interestingness as a quantitative *and* absolute measure, which for further studies does not require training data of any sort, drawing from the ascertainment of its strict relationship to the field of probability theory [3] and cognitive psychology.
Recent findings in the latter confirm that surprise is summoned by unexpected (schema-discrepant) events and its intensity is determined by the degree of schema-discrepancy [9], while the limits of hedonic analysis are highlighted, by mentioning for instance that the facial expression of surprise postulated by evolutionary emotion psychologists has indeed been found to rarely occur in real surprise[*ibid.*]. In "A cybernetic approach to aesthetics" [10] art is treated under the lens of data processing, according to which an information stream much higher or lower than 16 bits induces in the user respectively confusion and boredom, whereas close to that very threshold interestingness is considered to arise. From a psycholinguistic point of view other principles concerning the literary artefact come into consideration, such as the Uniform Information Density (UID): "speakers should plan their utterances so that elements with high information are lengthened, and elements with low information are shortened, making the amount of information transmitted per time more uniform (hence closer to the optimum)" [11].

3. The Lindenmeyer system and the Thue-Morse sequence

The Lindenmeyer system is a logical formalism originally employed as the basis of an axiomatic theory of biological development [12]. It consists of a quadruple (V, S, ω, P) where V is the set of variables and S the set of constants. The letters from V and S constitute the alphabet of the L system; ω is referred to as *start word* or *axiom* of the L-system; P is a set of ordered pairs of words over the alphabet, obtained by using the composition rules. Famous examples of L-modelable sequences are the Hilbert curve, in the scope of geometrical shapes, and the Thue-Morse sequence, in numbers. Just like in a magic squares, the sum of the elements of each row equals the sum of the elements of each column (4).
As observable in [4], the rows follow always one of two patterns, labelled by the letters A and B on the right edge. We notice that not only the single rows are palindrome, but also the letters column. Arranging a sequence in a matrix fashion allows to remark how patterns propagates also on a structural level, beside the sequential one[8]. In [3], they are displayed alongside with a L-system-inspired model for folktales: the first line introduces all the elements and sets necessary for the construction of the model: constants (C), narratemes (P), narratemes sequences (F), start-narrateme (α), end-narrateme (ω) and compositional rules (Z). The letters *d, k* and *l* stand for "deíxis", "katastrophé" and "lýsis", representing the key moment of the development of a story according to Van Dijk's narratological model [13][9]. In our formalization,

[8]This special distribution of the elements achieves the fulfilment of the UID theory, beyond the sentence, also at the text-level, in one of the most elegant possible ways.
[9]Temporally, they define the beginning, the middle, and the end, whereas narratologically they hint respectively to the introduction (world creation, initial situations description, characters presentation, etc.), the deflagration of conflicts and problems, and finally their solution and ending. For sake of computation, they will be rendered respectively as 1, 2, and 3.

$$
\begin{aligned}
E &= (C, P, F, \alpha, \omega, Z)\ ; \\
C &= (d, k, l)\ ; \\
C^1 &= (1, 2, 3); \\
C &\models C^1; \\
P &= \bigcup_{i=1}^{size(C)^{size(C)}} p_i\ ; \\
p &\in P = ccc,\ c \in C^1; \\
F &= vec(p_1 \xrightarrow{z \in Z} p_2 \xrightarrow{z \in Z} \dots \xrightarrow{z \in Z} p_n); \\
f &\in F = subvec(F); \\
\alpha &= [ddd] = [111]\ ; \\
\omega &= [lll] = [333]; \\
Z &: \\
[c*1] &\rightarrow [c*(2 \vee 3)]; \\
[c*2] &\rightarrow [c*3]; \\
[c*3] &\rightarrow [c*(c+1)*1] \vee \omega.
\end{aligned}
$$

a)

$$
\begin{aligned}
V &\rightarrow \quad 0, 1\ ; \\
\omega &\rightarrow \quad 0\ ; \\
P &\rightarrow \quad (0 \rightarrow 01), (1 \rightarrow 10).
\end{aligned}
$$

b)

$$
\begin{aligned}
V &\rightarrow \quad A, B; \\
S &\rightarrow \quad f, l, r; \\
\omega &\rightarrow \quad A; \\
P &\downarrow \\
A &\rightarrow \quad lBfrAfArfBl; \\
B &\rightarrow \quad rAflBfBlfAr\ .
\end{aligned}
$$

c)

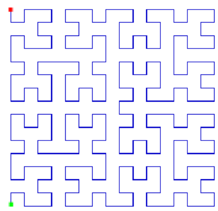

Figure 1: On the left, the typical folktale (as in Propp [2]) modelled as a L-system. On the right side: a) L-system for the Thue-Morse sequence; b) L-system for the Hilbert's curve; c) the Hilbert's curve.

a narratame is obtained by a combination of the above mentioned atomic constructors [10]. All possible narratemes are contained in the list P, obtained by performing a permutation with repetition on the given atomic constructors (precisely, the size of the set elevated to itself). The *alpha* and *omega* are the the required start- and end- constants. According to the model, after that the last number is upgraded, the one on its left is upgraded by one. If this is not possible, it means that the tale cannot be developed further. Every vector of function can be repeated as many time as desired.

4. The story as a Magic Square

The magic square is a square matrix, where the sum of the elements in every row, column and diagonal equals to $\frac{n*(n^2+1)}{2}$, whereby the *n* is the order of the matrix (the amount of elements for every row/column). The value of every position is represented by $x = M(i, j), i \wedge j \leq n$, where *i* represents the row and *j* the column. In this mathematical object, position and value are deeply entangled. In a square matrix of even order, the rule holds:

$$\forall i \in M \wedge \forall x \in X \Rightarrow M[i] = x \wedge M[x] = i$$

[10] For instance, in case the narrateme *Transfiguration* occurs in the canonical position (as in the list appearing in [8]), then it would be the Katastrophé - part of the Lysis - part of the Lysis-part of the story (hence, *llk*).

according to which the value in a given position equals the position where the number of its position is to be found (i.e., if in position 3 we find a 7, in the position 7 we shall find a 3). This profound relation between *value* and *position* is very similar to the one occurring in every fairytale and myths, as proven by Lévi-Strauss [14], according to which the mythical narration is both diacronic as syncronic: after its decomposition in *mythemes*, it can be arranged in a matrix where distinct *bundles of relationships* are highlighted by the columns [ibid.].

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & B \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & A \end{bmatrix} \begin{bmatrix} 6 & 32 & 3 & 34 & 35 & 1 \\ 19 & 7 & 11 & 27 & 28 & 8 \\ 30 & 14 & 16 & 15 & 23 & 24 \\ 19 & 20 & 22 & 21 & 17 & 18 \\ 25 & 29 & 9 & 10 & 26 & 12 \\ 36 & 2 & 33 & 4 & 5 & 31 \end{bmatrix} \begin{bmatrix} 12 & 14 & 15 & 16 & 17 & 17 \\ 11 & 7 & 9 & 10 & 10 & 12 \\ 1 & 4 & 2 & 3 & 5 & 6 \\ 5 & 2 & 4 & 3 & 1 & 0 \\ 7 & 11 & 9 & 8 & 6 & 6 \\ 18 & 16 & 15 & 14 & 13 & 13 \end{bmatrix}$$

The first matrix from the left, the Thue-Morse sequence, is an imperfect magic square, because it is composed only by binary numbers. However, we notice that also on the diagonals, as well as by rows and columns, the elements add up to 4[11]. In the middle, we find the so-called magic square of the sun (6x6). As easily observable, by rotating or transposing it, its properties do not change. Since in our framework every position can have only one value (the tale cannot be rotated or transposed), I have created a third matrix, where for every position (i,j) the value equals to $|n/2 - M(i,j)|$. Therein the "equal sums" property is held only for columns.

5. Proof of concept

Whereas the relaxed L-modelling of the fairy-tale according to Proppian rules has been an exclusively demonstrative logics exercise, whose real case scenario application is yet to be found[12], the above sketched hypothesis can be indeed tested. The samples selected for this proof of concept are: the time series of the 6x6 magic square (as for the third matrix in 4); the time series of the emotion intensity values of "Harry Potter and the Deathly Hallows" [8], transformed in order to obtain only absolute values; the time series of the self-information values of the Grimm's folktale "The Queen bee", obtained by combining for each sentence the Dependency Distance [15] (as measure for complexity) and the amount of Proppian narratemes occurring therein (as measure of *eventfulness* as in [5])[13]. The three vectors have been analysed by means of some of the aforementioned test statistics.

In the following, the Kolmogorov-Smirnov (KS) and the Epps-Singleton (ES) test statistics performed on the vectors 1 and 2 ('Magic Square' and 'Harry Potter'):

[11] This represents the *trait-d'-union* between Magic Square and Thue-Morse sequence.

[12] The fact that a fairy-tale can be framed by a recursive model would demonstrate that it is self-similar, as the magic square is.

[13] Further analysis conducted on a small corpus of German fairy-tales can be retrieved at https://github.com/Glottocrisio/GrimmHurst. The Hurst Exponent is there used as measure for the assessment of their degree of self-similarity according to various linguistic levels.

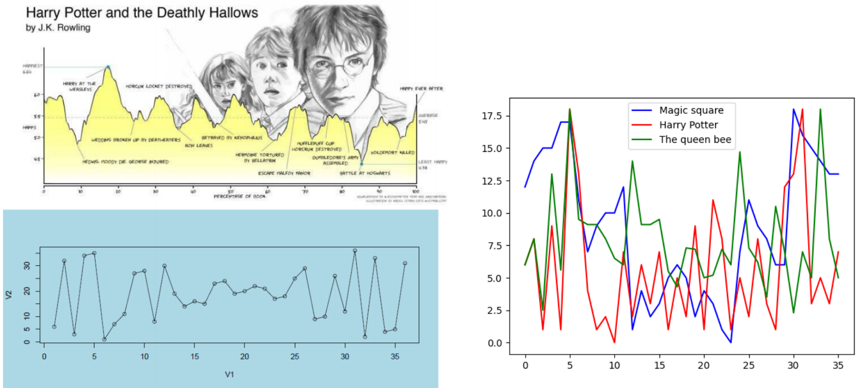

Figure 2: On the left: the emotion intensity distribution in the book "Harry Potter and the Deadly Hollows" as in [8] followed by the distribution of the 6x6 magic square displayed as a single sequence. On the right: a synoptic view of the distribution of information values for "Harry Potter" (red), "The queen bee" (green), and the magic square of the sun (blue).

```
KstestResult(statistic=0.3055, pvalue=0.0689)
Epps_Singleton_2sampResult(statistic=10.4553, pvalue=0.0334)
```

The KS and the ES tests performed on the vectors 1 and 3 ('Magic Square' and 'The Queen Bee'):

```
KstestResult(statistic=0.3055, pvalue=0.0684)
Epps_Singleton_2sampResult(statistic=17.2947, pvalue=0.0017)
```

The KS and the ES tests performed on the vectors 2 and 3 ('Harry Potter' and 'The Queen Bee'):

```
KstestResult(statistic=0.3888, pvalue=0.0081)
Epps_Singleton_2sampResult(statistic=18.1595, pvalue=0.0011)
```

6. Final considerations

The performed experiment is nothing more than a sketch of how a thorough analysis could be approached. The vectors are too few, too short and too unevenly obtained to represent a relevant dataset on which a reliable analysis can be performed. The present deepening on the interestingness does not pledge for its exclusivity, or even absolute prominence, in comparison to all other aspects playing an important role in narrative production and fruition [16]. The goal of this proposal and the application sketch is the initiation of a discussion about the re-framing of computational narratology in its needs, objectives and *rationalia*, triggered by the consideration of the overwhelming presence of approaches, that nowadays excessively push for rule- or probability- based automatic generation, whereby the aspects of interestingness, aesthetic pleasure and beauty, are often neglected for machine-friendlier metrics, such as coherence and correctness [17]. I have suggested that a story, as a magic square, is composed by values, that need to be distributed in a particular way in order to elicit aesthetic pleasure, yet maintaining the illusion of high entropy. As an L-system, the story can infinitely fill a finite space, as suggested by V. Propp in his work: every episode is a story made up of episodes.

References

[1] L. Itti, P. Baldi, Bayesian surprise attracts human attention, in: Advances in neural information processing systems, 2006, pp. 547–554.

[2] V. Propp, Morphology of the Folktale, 2nd. ed., University of Texas Press, Austin, 1968.

[3] M. Derezinski, K. R., An information theoretic approach to quantifying text interestingness (2014).

[4] R. J. Hilderman, H. J. Hamilton, Knowledge Discovery and Interestingness Measures: A Survey, Computer Science (1999) 28.

[5] H. O. Hatzel, E. Gius, H. Stiemer, C. Biemann, Narrativität und Handlung (2023).

[6] J. Gao, P. Pantel, M. Gamon, X. He, L. Deng, Modeling Interestingness with Deep Neural Networks, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 2–13. URL: http://aclweb.org/anthology/D14-1002. doi:10.3115/v1/D14-1002.

[7] M. Gamon, A. Mukherjee, P. Pantel, Predicting Interesting Things in Text (2014).

[8] Reagan A.J., Mitchell L., Kiley D., Danforth C. M., Dodds P. S., The emotional arcs of stories are dominated by six basic shapes (2016) 31.

[9] R. Reisenzein, G. Horstmann, A. Schützwohl, The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence, Topics in Cognitive Science 11 (2019) 50–74. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12292. doi:10.1111/tops.12292.

[10] H. W. Franke, A Cybernetic Approach to Aesthetics, Leonardo 10 (1977) 203. URL: https://www.jstor.org/stable/1573423?origin=crossref. doi:10.2307/1573423.

[11] A. F. Frank, T. F. Jaeger, Speaking rationally: Uniform information density as an optimal strategy for language production, in: Proceedings of the Annual Meeting of the Cognitive Science Society, Springer, 2008, pp. 939–944.

[12] A. Lindenmayer, Mathematical models for cellular interactions in development ii. simple and branching filaments with two-sided inputs, Journal of Theoretical Biology (1968) 300–315.

[13] T. A. van Dijk, Textwissenschaft. Eine interdisziplinäre Einführung, Niemeyer, 1980.

[14] C. Lévi-Strauss, Structural Anthropology, Basic Books, New York, 1963.

[15] H. Liu, Dependency distance as a metric of language comprehension difficulty (2008) 159–191.

[16] A. Zarcone, M. van Schijndel, J. Vogels, V. Demberg, Salience and Attention in Surprisal-Based Accounts of Language Processing, Front Psychol. (2016). doi:10.3389/fpsyg.2016.00844.

[17] L. Viola, Dexter: A post-authentic approach to heritage visualisation, in: Digital Humanities 2022 - Conference Abstracts, The University of Tokyo, 2022, pp. 547–554.