Understanding Narrative Interest: Some Evidence on the Role of Unexpectedness

Adrian Dimulescu, Jean-Louis Dessalles
Telecom ParisTech
{adrian.dimulescu, jean-louis.dessalles}@telecom-paristech.fr

Abstract

This study is an attempt to measure the variations of interest aroused by conversational narratives when definite dimensions of the reported events are manipulated. The results are compared with the predictions of the Complexity Drop Theory, which states that events are more interesting when they appear simpler, in the Kolmogorov sense, than anticipated.

Conversational narratives represent a significant part of spontaneous language, maybe up to 40% (Eggins & Slade, 1997, p.144). A significant part of individuals' social construction depends on their ability to tell interesting stories about events drawn from their daily life (Polanyi, 1979, Eggins & Slade, 1997, p.16, Norrick, 2000, p.84). The study presented in this paper is an attempt to measure the influence on interest of definite dimensions of the reported events.

Most studies on conversational narratives have been focused on sociolinguistic issues (e.g. Labov & Waletzky, 1967; Labov, 1997; Tannen, 1984). We concentrate here on cognitive aspects by investigating which factors best predict whether a narrative will be perceived as interesting. We designed an experiment in which participants were given two options and had to decide which one they considered more interesting (Table 1). For instance, in story 1, participants had to decide whether the stranger they encountered in the street asked the time or gave them a slap, and which alternative makes the better narrative. We compared results with the predictions of the Complexity Drop Theory, which states events are more interesting when they appear simpler, in the Kolmogorov sense, than anticipated.

Theoretical Background

The factors of interest in real life narratives have not been directly investigated until recently. Studies in related domains, especially episodic memory, showed that best remembered situations are atypical (Woll & Graesser, 1982; Shapiro & Fox, 2002), are inconsistent or deviate from norm (Stangor & Mcmillan, 1992). A single property, unexpectedness, subsumes all these characteristics. The main aim of the present paper is to show that unexpectedness determines narrative interest as well. The other major determining factor of interest, emotional intensity (Rimé, 2005), will not be considered here.

In previous work, we modeled unexpectedness as a complexity drop (Dessalles 2008a, 2008b). By analogy with the formal definition of Kolmogorov complexity, the *cognitive complexity* of a situation is defined as *the length of its shortest description available to the subject.* This definition makes correct predictions about some important aspects of perception (Chater, 1999). Complexity drop theory (CDT) (Dessalles 2008a, 2008b)[1] suggests that cognitive complexity is involved in some high-level cognitive processes as well, including narrative interest. The main claim of CDT is that people find a situation interesting when its representation is simpler than expected from the known workings of the world.

$$U(s) = C_w(s) - C(s)$$

Generation complexity $C_w(s)$ involves the complexity of all parameters that must be set for the "world machine" *i.e.* the "world" as we know it, to *generate* situation s. $C(s)$ is the minimal amount of information needed to *unambiguously describe* the situation. The difference $U(s)$ measures unexpectedness. For instance, CDT has been shown to correctly predict all dimensions of coincidences (Dessalles, 2008a). When two analogous situations s_1 and s_2 occur independently, their generation complexity $C_w(s_1\&s_2)$ is close to $2 \times C(s_1)$, whereas their description requires only $C(s_1) + C(s_2|s_1)$, which is smaller if the analogy is close. Individuals experience such situations as coincidences and consider them worth telling. CDT's scope includes other important aspects such as a new approach to subjective probability (Dessalles, 2008b).

Most situations are not unexpected, which means that their generation and their description require approximately the same amount of information. There are exceptions, however, and these exceptions make the topic of conversational narratives.

Experiment

We conducted an experiment to test CDT's predictions on interest. A corpus of 18 stories in French was established, using material from personal recorded sources and from a French community Web site (viedemerde.fr) where people describe short every-day life events, in an informal style, sometimes with a humorous touch. The stories were presented to 95 participants. The 14 most illustrative stories are listed below in their English version. Answers to the stories may seem *obvious*, as individuals have a clear intuition of what contributes to interest. What is less obvious is that the rich phenomenology of interest illustrated by the stories can be backed by a unified theory. The situation parallels the problem of syntax: people have a clear intuition of which sentences are syntactically correct in their mother language, but designing a predictive model of syntactic correctness remains a scientific challenge. Besides, experiment data offered some surprises, as explained in the discussion.

[1] See also www.unexpectedness.eu

1. *I was walking quietly in the street when a total stranger stops before me, looks at me and* [...] *before continuing his walk.* ⓐ gives me a phenomenal slap; ⓑ gives me a slap; ⓒ asks me the time.
2. *This afternoon, the police of Antibes discovered in the Baie des Anges area the floating dead bodies of two elegantly-dressed women. Apparently, the two accidents happened almost at the same time. Moreover, both women were wearing* [...]. ⓐ a red tattoo on the right arm representing the Tsuba-Kasai dragon; ⓑ a red tattoo on the right arm; ⓒ a tattoo on the right arm.
3. *I'd just bought a small Peugeot 106 ColorLine for 2000 euros. I had tried it the day before and it was very good. I turned the key, I started, I left the property of the former owner of the car when, coming from the left without looking, another* [...] *crashed into me.* ⓐ Peugeot 106 ColorLine; ⓑ Peugeot 106; ⓒ Peugeot.
4. *For the birthday of my little sister, my parents got her an 82-cm black flat TV screen. A little while ago, for my own, they got me* [...] ⓐ a 82-cm black baseball bat; ⓑ a black baseball bat; ⓒ a baseball bat.
5. *Wednesday, the city of Amiens police seized* [...] *kg of heroin at number 13 rue Fafet.* ⓐ 10; ⓑ 5; ⓒ 2.
6. *For a year, I had been thinking of changing my mobile phone at SFR (mobile operator). I finally decided to do so even if I had to pay a part because I did not have enough Red Square Points. I bought the new phone at 13:00.* [...] *I got a message from SFR: "Change your mobile, SFR offers you 15 000 Red Square Points."* ⓐ At 13:10; ⓑ At 14:00; ⓒ Two weeks later.
7. *I was lying on the ground under the trees on an autumn afternoon, when a leaf fell exactly* [...]. ⓐ on my nose; ⓑ on my face; ⓒ on my body.
8. *Two weeks after my car had been stolen, the police informed me that a car that might be mine was for sale on the Internet. They showed me the ad. The phone number had been identified. It was the mobile phone number of* [...]. ⓐ my office colleague; ⓑ a colleague of my brother's; ⓒ someone of my neighbourhood.
9. *I was walking in a street in downtown Paris when I heard someone calling me: it was a guy who I'd been babysitting* [...] *when he was a child. We exchanged addresses, it was really nice.* ⓐ for 2 years; ⓑ for 2 months; ⓒ a few evenings.
10. *It's funny, I found this on the Internet: the town of St-Chéron has* [...] *inhabitants.* ⓐ 4444; ⓑ 4000; ⓒ 3856.
11. *I got the license plate of my new car; I looked at the number I got, it was* [...]. ⓐ 999 NNN 91; ⓑ 253 NNN 91; ⓒ 253 UPV 91.
12. *In front of the coffee machine, I was talking to a colleague about SpongeBob and his best friend Patrick, the sea star. At some point I said: "I don't like Patrick, he's too stupid". At this precise moment, my boss passed by, his name is* [...] ⓐ Patrick Star; ⓑ Patrick; ⓒ Christopher.
13. *You know what? My neighbor, downstairs, she gave birth to* [...] *a few days ago. It was a premature birth. She said she was surprised to see them so small.* ⓐ quadruplets; ⓑ triplets; ⓒ twins.
14. *I had to pay a one year old fine of 400 euros because of the Treasury, which didn't send reminders to the correct address. When the sum was debited, there was exactly* [...] *euros on my account.* ⓐ 400; ⓑ 401; ⓒ 419.

All stories were presented in random order to all participants. Most participants were engineers and students working in fields other than cognitive science. For each story we isolated a parameter that we considered important to the relevance of the story and defined three options that would gradually affect the overall interest. One given participant only got to mark a preference between two such options, so that the overall ranking purpose of the test be less transparent. The test was implemented as a Web application. Participants had to fill-in a missing excerpt by clicking on one of two randomly chosen options that were displayed below the main text. Participants had "to select the option that made the story more interesting". A reward (USB stick) was to be given to those whose overall choices were most consistent with the majority. Results are listed in Table 1.

Ranking

At the end of the process, each story produced a list of three paired comparisons between its fill-in options. We computed a ranking of story versions to see if it was congruent with the predictions of CDT. In a manner similar to (Saaty, 1994), we want to calculate the rank r_i of a version i, given pairwise relations between versions. Knowing the win/lose history of competition between version i and j, we define w_{ij} as the ratio $wins(i)/wins(j)$. In order to avoid division by zero, we accorded one vote by default to all versions. For example, if the comparison between two versions of a given story was presented to 30 participants of which 23 chose the first version while the remaining 7 participants chose the second, then $w_{ij} = 24/8$. Note that $w_{ji} = 1/w_{ij}$. We want a story version that performed well against other versions to have a rank that reflects its win, taking into account both the rank of the defeated version and the bluntness of the win. A possible calculation is $r_i = k\sum_j w_{ij} r_j$, rewritten as $\frac{1}{k}R = WR$ which amounts to finding an eigenvector of the symmetrically reciprocal matrix W (Farkas, 2007). Ranks are then normalized so that each eigenvector sum to 100.

Table 1 shows the overall results and rankings. For each story, versions are ordered from most interesting to least interesting according to CDT. Most of the comparisons (those marked with *) were found statistically significant ($p < .05$) on a standard binomial test targetted at detecting a marked preference for one particular version. Participants' choices are highly congruent with the predictions. Rankings show that participants altogether never preferred option (c), which is excluded by the model. For 17 stories out of 18, option

Table 1: Comparison between paired options and the resulting option rankings for each story

	comp.	rank		comp.	rank
	a-b 22/18	a 53		a-b* 35/1	a 92
1	b-c* 25/4	b 41	10	b-c* 21/11	b 4
	a-c* 27/3	c 6		a-c* 30/2	c 3
	a-b 17/14	a 40		a-b* 28/5	a 76
2	b-c* 34/5	b 49	11	b-c* 22/11	b 15
	a-c* 23/9	c 11		a-c* 30/4	c 9
	a-b* 24/8	a 61		a-b 18/16	a 52
3	b-c* 32/4	b 32	12	b-c* 29/4	b 43
	a-c* 28/5	c 7		a-c* 31/3	c 5
	a-b 18/16	a 38		a-b* 24/11	a 56
4	b-c* 22/11	b 39	13	b-c 21/12	b 27
	a-c* 20/14	c 23		a-c* 25/8	c 17
	a-b* 28/13	a 54		a-b 13/20	a 33
5	b-c 20/11	b 28	14	b-c* 29/5	b 56
	a-c* 21/8	c 18		a-c* 24/9	c 11
	a-b* 29/4	a 79		a-b* 24/9	a 56
6	b-c* 28/6	b 16	15	b-c* 28/6	b 33
	a-c* 31/3	c 5		a-c* 26/8	c 11
	a-b* 26/4	a 68		a-b* 30/7	a 70
7	b-c* 36/4	b 25	16	b-c 17/15	b 16
	a-c* 24/6	c 7		a-c* 25/5	c 14
	a-b* 25/6	a 65		a-b* 31/12	a 60
8	b-c* 24/10	b 22	17	b-c* 21/8	b 28
	a-c* 26/7	c 13		a-c* 24/6	c 12
	a-b* 25/11	a 58		a-b* 32/5	a 66
9	b-c 20/12	b 26	18	b-c* 30/1	b 31
	a-c* 25/7	c 16		a-c* 28/4	c 3

(a), which is the most congruent with the model, is significantly ranked best. In only one story, option (b) tended to be preferred to option (a) (see discussion).

Analysis

Stories are meant to test various parameters that influence interest: coincidences (stories 2, 3, 4, 12), quantitative deviation (5, 13), qualitative deviation (1), temporal (6, 17), spatial (7) and conceptual (8, 15, 16, 14) proximity, fortuitous encounters (9) and structure (10, 11). Though probability-based theories provide approximations for quantitative deviation and proximity (Dessalles 2008b), no current theory can account for all of these influences. CDT makes predictions in each case, and they turn out to be generally confirmed by our test.

Coincidences

CDT predicts analogies to be interesting because the "world machine" must generate the two terms of the coincidence separately, whereas the "observation machine" can use one situation to describe the other (Dessalles, 2008a). Unexpectedness amounts to $U(s_1|s_2) = C_w(s_1) + C_w(s_2) - C(s_1) - C(s_2|s_1)$. If s_1, alone, is not unexpected, then $C_w(s_1) = C(s_1)$, and:

$$U = C_w(s_2) - C(s_2|s_1)$$

Stories 2, 3 and 12 in the corpus involved a coincidence(story 4 was also intended to present a coincidence, but see discussion). Participants were highly sensitive to it, as they very significantly rejected option (c) in each case. For instance, in story 3, there is an accident between two nearly identical cars. Participants preferred the second car to be of the same series as the first one (Peugeot/106/ColorLine) over it being merely of the same type (Peugeot/106), and they preferred the latter over of the second car being of the same make (Peugeot).

These preferences are consistent with CDT. To compute $C_w(s_2)$, we may consider that the "world machine" had to "choose" among N cars to pick the one involved in the accident. We thus have $C_w(s_2) = \log_2 N$. The computation of $C(s_2|s_1)$ can use the common feature f, which is available with no complexity from s_1. If the number of cars with feature f is n_f, then one needs $\log_2 n_f$ bits to discriminate the actual car. Unexpectedness thus amounts to:

$$U(s_1 \& s_2) = \log_2 N - \log_2 n_f$$

The prediction is that the most specific common feature f will be considered most interesting. This is indeed what we observed.

Quantitative Deviation

According to CDT, situations that are unique (or easy to single out) for a *simple* reason will be unexpected when they occur. Atypical situations are interesting because they are easy to single out. For instance, in story 5, participants highly significantly preferred option (a) (10 kilos of heroin) over option (b) (5 kilos), and strongly preferred the latter over (c) (2 kilos).

A situation s is considered atypical if it departs from a typical reference r by k standard deviations along feature f. If the "world machine" is considered equivalent to a lottery, then $C_w(s|r) = \log_2 N$, where N is the number of situations corresponding to r. The complexity involved in describing s using r and f is at most: $C(s) = C(r) + C(f|r) + C(s|r\&f)$. We will suppose that $C(f|r) = C(f)$. To compute $C(s|r\&f)$, one may rank situations in r by values of f with negligible complexity (note that complexity refers to the size of algorithms, not to their execution time). If s is extreme, it will appear among the firsts in this ranking. At any rate, feature f distinguishes a smaller number of situations n_f inside the total N. We may then write $C(s|r\&f) = \log_2 n_f = \log_2 pN$, with $p = n_f/N$. If $A = -\log_2 p$ is an expression of the atypicality of s, then $C(s|r\&f) = \log_2 N - A(k)$, where function A only depends on the statistical distribution of r along f. For a Laplace-Gauss distribution, $A(k) \sim k^2$. If r is not unexpected, we eventually get[2]:

[2] See details on www.unexpectedness.eu/Fish.html

$$U(s) \approx A(k) - C(f)$$

If participants perceive the two options proposed to them in story 5 as involving different values of k, the prediction is that they will opt for the larger one. Our results clearly support this prediction.

Qualitative Deviation

Some events happen, though they were previously thought to be nearly impossible. This property merely means that the complexity $C_w(s)$ required to generate such an event s is large. It amounts to $C(H)$, where H is the simplest causal scenario that explains s given the workings of the known world (see discussion). According to CDT, unexpectedness is $U(s) = C(H) - C(s)$.

For instance, in story 1, explaining why someone asked the time (option (c)) may be as simple as "because he forgot his watch", whereas an explanation of an unmotivated slap would require quite a bigger scenario, and consequently a much larger $C(H)$. On the other hand, complexity $C(s)$ may be computed through a reference frame r and a feature f of s, as previously: $C(s) = C(r) + C(f) + C(s|r\&f)$. In story 1, r corresponds to a typical scene in the street. For options (a) and (b), f corresponds to the fact of being slapped by a stranger. If such a description can be regarded as capturing a unique situation, then $C(s|r\&f) = 0$. Unexpectedness therefore amounts to:

$$U(s) = C(H) - C(r) - C(f)$$

For option (c) (asking the time), $C(s|r\&f)$ has a significant value, since scenes featuring one person asking the time in the street are numerous and thus difficult to discriminate. Unexpectedness will thus be high in options (a) and (b) and close to zero in (c).

Experiment with story 1 confirms the prediction. Note that option (a) (phenomenal slap) and (b) (slap) are not significantly distinguished in participants' preferences. This is consistent with the fact that the two corresponding causal scenarios are of similar complexity.

Proximity

Mentioning a reference point (spatial, temporal, social or other) makes things that happened in its proximity worth telling. Proximity can be thought of as a relaxation of coincidence. In coincidences, $C(s_2|s_1)$ is lowered by identical features shared by s_1 and s_2. Proximity effects also diminish the amount of information needed to define s_2 from s_1, by using the latter as a reference point from which it is easy to locate s_2. CDT thus predicts that events happening closer to each other than usual will add to the interest.

Seven stories in our experiment involve proximity effects. In story 6, promotional offers by the mobile phone operator are expected to occur, say, once every two months. The "world machine" requires $C_w(t_2) = \log_2\ 86400 = 16.4$ bits to locate the specific minute t_2 when the offer arrives. For the "observation machine", however, the complexity of t_2 is only $C(t_2|t_1) = \log_2\ 10 = 3.3$ bits, as it can use the moment t_1 when the mobile has been bought as reference point. This contrast contributes 13 bits to unexpectedness (note that the value would be the same with a different time resolution).

More generally, if the event is expected to occur with time density D and is observed at temporal distance d, then CDT predicts the contribution to unexpectedness to be[3]:

$$U = -\log_2\ (Dd)$$

In a two-dimensional space, a factor of 2 must be included.

In story 6, the three options are unexpected by 13 bits, 10.5 bits and 2 bits respectively. Participants' preferences strongly confirm this effect, as (a) was favored over (b) (29/4) and over (c) (31/3), and (b) was favored over (c) (28/6). Story 7 shows a similar confirmation pattern, this time in space.

Story 8 illustrates the role of social proximity. Unexpectedness here relies on the fact that the thief happens to be simpler than expected. The generic unknown person P has a complexity $C_w(P)$ that we might estimate by the logarithm $\log_2\ N$ of the population in the region that the subject takes as reference, probably the city. If P happens to be someone living in a neighborhood of size n, then the contribution to unexpectedness is $U = \log_2\ N - \log_2\ n - C(d)$, where d is the concept of neighborhood. This extensional computation represents a last resort, as a computation through the graph of acquaintances often provides a smaller value for $C(P)$. In this case, the complexity of a node is the minimum information needed to reach the node of P. In story 8, the expression "my colleague" suggests that the complexity of P is zero once the concept of colleague is installed, as P comes first in the list. Results unambiguously confirm the prediction: option (a) (my colleague) is significantly preferred to (b) (a colleague of my brother), and both (a) and (b) are preferred to (c) (someone living in my neighborhood).

Fortuitous Encounters

Fortuitous encounters make good conversational stories. The unexpectedness they produce can be written as $C(l) - C(P)$, where l is the location of the encounter and P the person met (Dessalles, 2008a).

The options of story 9 vary $C(P)$ by changing the duration of the former acquaintance with P: two years (a), two months (b) and a few days (c). A longer period of acquaintance puts P higher in rank in the list of people that one knows personally. $C(P)$ may be assessed by the logarithm of the rank in that list. Answers to story 9 confirm the prediction. Note that the preference of (b) over (c) is weak, which may be explained by the fact the two options do not change the acquaintance with P significantly.

Structure

One of the most immediate predictions of CDT is that the occurrence of remarkable structures will increase interest. The

[3]See details on www.unexpectedness.eu/NextDoor.html

"world" requires the same efforts to generate a typical, unremarkable structure s_r and the actual one: $C_w(s) = C(s_r)$. But remarkable structures are simple and thus unexpected:

$$U(s) = C(s_r) - C(s)$$

Stories 10 and 11 in our corpus involved simple structures. In story 10 (a), number 4444 saves three instantiations, as the digit 4 is merely copied. Unexpectedness is then $U_a = 3 \times \log_2 10 = 10$ bits. In option (b) unexpectedness, for the same reason, amounts at least to $U_b = 6.6$ bits, whereas it is zero for option (c). Similar computations for story 11 give $U_a = 16$, $U_b = 9.4$ and $U_c = 0$ bits. Results confirm these predictions. In both stories, the simplest structure (a) was strongly preferred to (b) and (c). Option (b) was preferred to (c), though less significantly.

Discussion

Most outcomes of the experiment described in this paper are in full accordance with the predictions of CDT. There are a few discrepancies, though, that will be discussed below. Before, we must address the recurrent question of whether complexity can be measured.

Measuring Complexity

A common claim, easy to prove, is that Kolmogorov complexity is not computable. We avoid this difficulty by considering two restrictions. First, we define complexity as the length of the most concise description currently available, and not as the objective minimum. This may lead to an underestimate of unexpectedness and explain why some subjects may fail to capture the interest that other subjects find in a story.

Our second restriction is that cognitive complexity is computed on a specific machine, the "cognitive machine", *i.e.* the cognitive tools considered available to humans. This presupposes that the cognitive model we use be made explicit in each case. In the examples previously commented on, most computations were made using minimal assumptions about cognitive abilities. As we showed, the computation of cognitive complexity is perfectly tractable in most cases. Three aspects of complexity are, however, external to the model: conceptual complexity, structural complexity and causal complexity.

Conceptual complexity is needed when prototypes and features, noted r and f in our examples, are involved. Though we did not explore it yet, a way to assess $C(r)$ or $C(f)$ would be to use minimal path length in ontology graphs.

Structural complexity can be approximated using common compression programs, such as `gzip`, `bzip2` (Cilibrasi & Vitanyi, 2005). For each option of story 11, we measured `bzip2` output size (on input replicated 1000 times to compensate for compressor inadequacy for small strings). We got, in bytes: (a) 64, (b) 71 and (c) 73. If (c) is taken as the expected reference, we get in bits: $U_a = 9, U_b = 2$. Though these values are underestimates (compare to the values computed above for the same story: $U_a = 16$ and $U_b = 9.4$), they preserve hierarchy and may be useful when complexity must be assessed by an automated system.

Causal complexity $C_w(H)$ is equivalent to explanation parsimony (Feldman, 2004; Chaitin, 2004). The complexity of a causal scenario depends on the subject's capacity to imagine plausible causes through abduction. Generation complexity $C_w(s)$ is then recursively computed from the generation complexity of these causes (2008b). Since modeling abduction is problematic when common sense is involved (Magnani, 2001), we are in search of an indirect method. A promising technique consists in using an information distance such as the normalized Google distance (NGD) (Cilibrasi & Vitányi, 2007). The Web offers the textual trace of countless non-unexpected events. Statistical co-occurrence of words in texts is thus expected to correlate negatively with the complexity $C_w(s)$ of a situation s described using these words.

In story 1, the occurrence of a man slapping another in the street is supposed to be more difficult to generate than if the man merely asks the time. Using a straightforward application of the NGD formula, we calculated the distance (normalized between 0 and 1) from "street" to "slap" (0.74); it is significantly larger than the distance from "street" to "ask" (0.15). By comparison, "street" is close to "building" (0.08) and far from an abstract concept like "configure" (0.84). These distances are claimed to offer reliable estimates of complexity (Cilibrasi & Vitányi, 2007), although some effort is still needed to render NGD-type calculations more robust across search engines and text corpora (Lindsey, Veksler, Grintsvayg, & Gray, 2007) and less dependent on word morphology.

Prediction Accuracy

Most judgments on the test stories are in accordance with predictions of the CDT. Some results are, however, somewhat surprising.

- Stories 12 and 2: there is no preference for the most specific common characteristics over the second most specific one. For instance, Peugeot/106/Colorline is not significantly preferred to Peugeot/106, contrary to what CDT predicts. Our explanation is that participants might have found the former too unrealistic to make a credible story. One participant's comment reads: "It's the second time I select something less interesting but more credible." Further investigation is needed to make this point clear.
- Story 4: the main point of the story is the contrast between presents; many participants apparently did not pay attention to coincidental characteristics (82 cm and black color), which turned out to be irrelevant to them.
- Story 14: this is the only real surprise of the experiment. Participants tended to prefer 401 to 400 Euros, while CDT predicts that the latter be simpler. A possible explanation is that 401 uses a frame/feature predicative representation (400 and 1) that makes it both simple and unique, whereas 400 refers to a generic round-number frame (as for 1001 nights *vs.* 1000 nights). We plan to test this hypothesis.

Conclusions

Discovering a single cognitive account for the human sensitivity to interest is a challenging objective. One may think of a variety of separate explanations for the different story types: one explanation for analogies, one for proximity, one for structure, and so on. The experiment described in this paper supports the idea that a single and simple principle: complexity drop, makes all correct predictions with no need to invoke *ad hoc* hypotheses. In particular, the feeling of improbability that is often mentioned when reading the stories of our test is not a separate dimension of interest. Subjective probability p can be deduced from unexpectedness U (Dessalles, 2008b) through:

$$p = 2^{-U}$$

According to CDT, an interesting event is required to be more complex to produce than to (unambiguously) describe. This requirement is quite difficult to match, and only relevant stories seem to pass the test. Note that a common misconception consists in considering the length of a verbal description as an indicator of description complexity. But if the description (e.g. the presence of a monstrous animal in my garden) turns an ordinary situation into a unique one, the net benefit in terms of unexpectedness may be significant. The prediction is, moreover, that only elements that contribute to unexpectedness will be mentioned in the verbal description.

The main difficulty with our experimental approach to interest is that stories in the test lie halfway between fiction and non-fiction. For instance, in story 13, 21 subjects chose “triplets” against “twins”, but 12 chose the latter. If participants were to believe that both events really happened, they would probably favor the newsworthiness of “triplets” more clearly. We are currently working at an experiment design in which judgments of interest are closer to real-life experience.

Another perspective is to make complexity computations automatic, using conceptual knowledge retrieved from ontologies like Wordnet, and assessing judgments of typicality from Web-based distances. Anticipated applications are the computation of newsworthiness and event-oriented search engines.

Acknowledgements

We are thankful to Yves Guiard for his advice.

References

Chaitin, G. (2004). On the intelligibility of the universe and the notions of simplicity, complexity and irreducibility. In Hogrebe & Bromand (Eds.), *Grenzen und grenzüberschreitungen* (pp. 517–534). Akademie Verlag.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology Section A*, *52*(2), 273–302.

Cilibrasi, R., & Vitanyi, P. M. B. (2005). Clustering by compression. *Information Theory, IEEE Transactions on Information Theory*, *51*(4), 1523–1545.

Cilibrasi, R., & Vitányi, P. M. B. (2007). The Google similarity distance. *Knowledge and Data Engineering, IEEE Transactions on Knowledge and Data Engineering*, *19*(3), 370–383.

Dessalles, J. L. (2008a). Coincidences and the encounter problem: A formal account. In *30th annual conference of the cognitive science society* (pp. 2134–2139).

Dessalles, J. L. (2008b). *La pertinence et ses origines cognitives*. Hermes-Science publications.

Eggins, S., & Slade, D. (1997). *Analysing casual conversation*. London: Equinox.

Farkas, A. (2007). The analysis of the principal eigenvector of pairwise comparison matrices. *Acta Polytechnica Hungarica*, *4*(2).

Feldman, J. (2004, October). How surprising is a simple pattern? Quantifying ”Eureka!”. *Cognition*, *93*(3), 199–224.

Labov, W. (1997). Some further steps in narrative analysis. *Journal of Narrative and Life History*, *7*(1-4), 395–415.

Labov, W., & Waletzky, J. (1967). *Narrative analysis: Oral versions of personal experience*. Seattle, WA: University of Washington Press.

Lindsey, R., Veksler, V. D., Grintsvayg, A., & Gray, W. D. (2007). Be wary of what your computer reads: The effects of corpus selection on measuring semantic relatedness. In *Proceedings of the 8th international conference on cognitive modeling*. Ann Arbor, MI.

Magnani, L. (2001). *Abduction, reason and science - processes of discovery and explanation*. New York: Kluwer Academic.

Norrick, N. R. (2000). *Conversational narrative: storytelling in everyday talk*. Amsterdam: John Benjamins Publishing Company.

Polanyi, L. (1979). So what’s the point? *Semiotica*, *25*(3), 207–241.

Rimé, B. (2005). *Le partage social des émotions*. Paris: PUF.

Saaty, T. L. (1994). How to make a decision: The analytic hierarchy process. *Interfaces*, *24*(6), 19–43.

Shapiro, M. A., & Fox, J. R. (2002). The role of typical and atypical events in story memory. *Human Communication Research*, *28*(1), 109–135.

Stangor, C., & Mcmillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61.

Tannen, D. (1984). *Conversational style - analyzing talk among friends*. Norwood: Ablex Publishing Corporation.

Woll, S. B., & Graesser, A. C. (1982). Memory discrimination for information typical and atypical of person schemata. *Social Cognition*, *1*, 287–310.