

Riverrun3. Propuestas desde la perspectiva
científica.

Ángel M. Núñez

November 14, 2014

Estas notas son sólo para dejar constancia de lo pensado y compartirlo. Más adelante las podemos incluir en un documento más general.

0.1 Vocabulario

Una medida cuantitativa propuesta para caracterizar la dinámica de manera global sería la “riqueza” del texto en forma de vocabulario $V(t)$ (entendido como número de palabras diferentes que componen el texto en cada instante). La evolución temporal se podría describir con una especie de “master equation”:

$$V(t+1) = V(t) + s(t) - d(t) \quad (1)$$

en la que habría que modelizar de alguna forma los términos fuente $s(t)$ de las palabras nuevas que los autores van introduciendo y sumidero $d(t)$ de las que van eliminando en sus sucesivas modificaciones.

Si consideramos que cada obra se compone no en saltos de tiempo discretos sino en tiempo real, podríamos formular de idéntica manera la ecuación en forma continua:

$$\frac{dV}{dt} = n_{s-d}(t) \quad (2)$$

en que la parte derecha sería un término fuente neto resultante del balance $n_{s-d}(t) = s(t) - d(t)$.

0.1.1 Tiempo de cristalización

Desconocemos los detalles de la evolución temporal de nuestra variable “vocabulario” $V(t)$ pero sabemos que inicialmente ha de crecer desde $V(t=0) = 0$ (principio del experimento) y que, transcurrido un tiempo característico t_c (todavía no caracterizado) ha de hacerse constante

$$\frac{dV}{dt}(t=t_c) = 0 \Rightarrow V(t_c) = cte, \quad (3)$$

pues nuestro sistema “cristaliza” y el texto no experimenta modificaciones posteriores (observación empírica). Nuestra conjetura es que este tiempo característico es una función del radio de visibilidad de cada autor r_v y del número total de autores N que participa en el experimento: $t_c = t_c(r_v, N)$.

0.2 Actividad

La variable vocabulario, pese a ser una medida de la “complejidad” del texto, no informa por sí sola de la actividad que puede estar teniendo lugar en la obra en un determinado instante ($V(t) = cte$ no implica que la obra esté cristalizada, simplemente que $s(t) = d(t)$). No sabemos si, antes de la cristalización, el sistema llega a un régimen estacionario en que el volumen de vocabulario no varía substancialmente pero se siguen haciendo modificaciones en el texto.

Habría pues que proponer otra medida que diera cuenta de las modificaciones que el texto está experimentando más allá del vocabulario: una suerte de “ $sum[diff(text_t2, text_t1)]$ ” entre instantes de tiempo sucesivos a la que podríamos llamar $a(t)$.

0.3 Perturbaciones/Cascadas de actividad

Una vez el sistema ha “cristalizado”, se puede intentar “perturbar” introduciendo en algún lugar del texto una palabra “provocadora” o “sugereante”, una idea nueva, etc. y medir la actividad a la que da lugar en forma de modificaciones que los autores introducen para adaptar la obra a la nueva situación.

$$A = \int_{t_{c1}}^{t_{c2}} a(t) dt \quad (4)$$

Medir la distribución de tamaños $P(A)$ de estas “cascadas de actividad” nos puede dar indicios de si el sistema evoluciona en las inmediaciones de un punto crítico ($P(A) \sim A^{-\alpha}$) y si se puede relacionar esto con el punto óptimo de un sistema para procesar información.

0.4 Trazabilidad. Evolución de trayectorias

Otra perspectiva que se puede adoptar para la caracterización del sistema es concebir el texto como un medio difusivo/dispersivo de ideas/palabras/etc. y observar las trayectorias que diferentes palabras/ideas experimentan a lo largo del texto: la velocidad con que se propagan, el tiempo que sobreviven hasta la extinción, la huella que dejan en el texto en forma de repeticiones...

Para que esta aproximación sea modelizable de forma satisfactoria, habría que definir adecuadamente que “unidades semánticas” habríamos de trazar y cómo medirlas: ¿Sustantivos, adjetivos y verbos? ¿Palabras y sus derivadas (género, número, adjetivación de sustantivos, sustantivación de verbos, etc.)? ¿Palabras y sus sinónimos?...

0.5 ¿“El autor promedio”?

¿Tiene sentido hablar del “autor promedio”? ¿Se mueven los autores durante el proceso de creación en un rango de parámetros próximo entre ellos? ¿Existe, p.e., un número característico $\langle c \rangle$ de caracteres por autor?:

$$\langle c \rangle = \frac{C}{N} = \frac{1}{N} \sum_i^N c_i \quad (5)$$

donde C es el número total de caracteres de la obra, N es el número total de autores y c_i es el número de caracteres que utiliza el autor i . ¿Y un número característico $\langle p \rangle$ de palabras por autor?:

$$\langle p \rangle = \frac{P}{N} = \frac{1}{N} \sum_i^N p_i \quad (6)$$

¿Tiene sentido hablar de un tamaño de vocabulario promedio $\langle v \rangle$ para cada autor en función del número de autores?:

$$\langle v \rangle = \frac{V}{N} \quad (7)$$

donde V es el tamaño de vocabulario presente en la obra. Esta variable $\langle v \rangle$ es diferente en su definición de las dos anteriores: pese a que V podría continuar siendo una variable extensiva ($V = V(N)$), su dependencia con el número de autores N dejaría de resultar trivialmente aditiva, dado que el vocabulario v_i de cada uno de los autores i tendrá palabras en común con el resto:

$$V \leq \sum_i^N v_i \quad (8)$$