Pergamon

*Mathl. Comput. Modelling* Vol. 19, No. 12, pp. 27–36, 1994

0895-7177/94 $7.00 + 0.00

0895-7177(94)00079-4

# Poisson Evolution in Word Selection

A. F. Badalamenti, R. Langs, G. Cramer and J. Robinson
The Nathan Kline Institute for Psychiatric Research
Building #37
Orangeburg, NY 10962, U.S.A.

*(Received and accepted February 1994)*

**Abstract**—This paper presents the finding that the invocation of new words in human language samples is governed by a slowly changing Poisson process. The time dependent rate constant for this process has the form

$$\lambda(t) = \lambda_1 (1 - \lambda_2 t) e^{-\lambda_2 t} + \lambda_3 (1 - \lambda_4 t) e^{-\lambda_4 t} + \lambda_5, \qquad \text{where } \lambda_i > 0, \quad i = 1, \dots, 5.$$

This form implies that there are opening, middle and final phases to the introduction of new words, each distinguished by a dominant rate constant, or equivalently, rate of decay. With the occasional exception of the phase transition from beginning to middle, the rate $\lambda(t)$ decays monotonically. Thus, $\lambda(t)$ quantifies how the penchant of humans to introduce new words declines with the progression of their narratives, written or spoken.

**Keywords**—Word analysis, Stochastic model, Evolutionary process, Rate constants, Poisson.

## INTRODUCTION

The present paper began with an investigation of how new words are introduced into written or spoken texts. Prior work [1] found that the introduction of new words is approximately Poisson with respect to word position into the narrative. The data for these prior works were text samples less than several hundred words long, in contrast to the present works each of which exceed 600 words in length.

More careful study of longer narratives suggests that the new word process is governed by a slowly evolving Poisson model, that is, a model for which the probability of an event in a small time interval $h$ has the form $\lambda(t)h + o(h)$ where $o(h)/h \to 0$ as $h \to 0$ [2–5]. Figure 1 gives the histogram for waiting times to new words at the end of each quarter of Martin Luther King's speech ("I had a dream ... "), where the waiting time is the number of words between a new word and a next new word.

Note that the sample standard deviations are comparable to the means, a sign of an exponential population. This suggests that the waiting times between new words have a signature of a Poisson process—a negative exponential density. The sample means, whose reciprocals should estimate the Poisson rate constant, grow from quarter to quarter. This is evidence for a Poisson process which evolves by slowing.

The repeated discovery of such findings as in Figure 1 across many samples led to the present position that a slowly evolving Poisson process governs the way speakers and writers introduce new words into their text.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

```
                                      SYMBOL   COUNT        MEAN       ST.DEV.
                                        X       205         2.000          1.785
                               EACH SYMBOL REPRESENTS        5 OBSERVATIONS
INTERVAL                                                    FREQUENCY  PERCENTAGE
NAME          25   50   75  100  125  150  175  200        INT.  CUM.  INT.  CUM.
          +----+----+----+----+----+----+----+----+
*1        +XXXXXXXXXXXXXXXXXXXXXX                          108   108   52.7  52.7
*2        +XXXXXXXXXXX                                      56   164   27.3  80.0
*3        +XXXX                                             21   185   10.2  90.2
*4        +X                                                 5   190    2.4  92.7
*5        +X                                                 3   193    1.5  94.1
*6        +X                                                 5   198    2.4  96.6
*7        +X                                                 4   202    2.0  98.5
*8        +                                                  0   202    0.0  98.5
*9        +                                                  2   204    1.0  99.5
*10       +                                                  0   204    0.0  99.5
```

(a). Histogram of new word waiting times, first quarter of Martin Luther King's speech.

```
                                      SYMBOL   COUNT        MEAN       ST.DEV.
                                        X       336         2.438          2.126
                               EACH SYMBOL REPRESENTS        5 OBSERVATIONS
INTERVAL                                                    FREQUENCY  PERCENTAGE
NAME          25   50   75  100  125  150  175  200        INT.  CUM.  INT.  CUM.
          +----+----+----+----+----+----+----+----+
*1        +XXXXXXXXXXXXXXXXXXXXXXXXXXXXX                   144   144   42.9  42.9
*2        +XXXXXXXXXXXXXXXXXXX                              94   238   28.0  70.8
*3        +XXXXXXXX                                         38   276   11.3  82.1
*4        +XXX                                              14   290    4.2  86.3
*5        +XX                                               12   302    3.6  89.9
*6        +XXX                                              15   317    4.5  94.3
*7        +X                                                 7   324    2.1  96.4
*8        +X                                                 4   328    1.2  97.6
*9        +X                                                 5   333    1.5  99.1
*10       +                                                  1   334    0.3  99.4
```

(b). Histogram of new word waiting times, first half of Martin Luther King's speech.

```
                                      SYMBOL   COUNT        MEAN       ST.DEV.
                                        X       433         2.841          2.683
                               EACH SYMBOL REPRESENTS        5 OBSERVATIONS
INTERVAL                                                    FREQUENCY  PERCENTAGE
NAME          25   50   75  100  125  150  175  200        INT.  CUM.  INT.  CUM.
          +----+----+----+----+----+----+----+----+
*1        +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX                161   161   37.2  37.2
*2        +XXXXXXXXXXXXXXXXXXXXXXX                         113   274   26.1  63.3
*3        +XXXXXXXXXXXX                                     61   335   14.1  77.4
*4        +XXXX                                             21   356    4.8  82.2
*5        +XXXX                                             20   376    4.6  86.8
*6        +XXXX                                             19   395    4.4  91.2
*7        +XXX                                              14   409    3.2  94.5
*8        +X                                                 6   415    1.4  95.8
*9        +X                                                 6   421    1.4  97.2
*10       +                                                  2   423    0.5  97.7
*11       +                                                  1   424    0.2  97.9
*12       +X                                                 3   427    0.7  98.6
*13       +X                                                 3   430    0.7  99.3
*14       +                                                  0   430    0.0  99.3
*15       +                                                  1   431    0.2  99.5
```

(c). Histogram of new word waiting times, first three quarters of Martin Luther King's speech.

Figure 1. **Note:** All histograms are plotted on the *same scale*. Frequencies of less than 5 do not appear.

```
                                    SYMBOL   COUNT        MEAN        ST.DEV.
                                      X       516         3.167        3.307
                              EACH SYMBOL REPRESENTS        5 OBSERVATIONS
INTERVAL                                                  FREQUENCY PERCENTAGE
NAME         25    50    75   100   125   150   175   200  INT.  CUM.  INT.  CUM.
          +----+----+----+----+----+----+----+----+
*1        +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX            180   180   34.9  34.9
*2        +XXXXXXXXXXXXXXXXXXXXXXXXXX                      131   311   25.4  60.3
*3        +XXXXXXXXXXXXXXX                                  74   385   14.3  74.6
*4        +XXXXX                                            26   411    5.0  79.7
*5        +XXXX                                             22   433    4.3  83.9
*6        +XXXXX                                            26   459    5.0  89.0
*7        +XXX                                              15   474    2.9  91.9
*8        +XX                                                9   483    1.7  93.6
*9        +XX                                               10   493    1.9  95.5
*10       +X                                                 3   496    0.6  96.1
*11       +                                                  1   497    0.2  96.3
*12       +X                                                 5   502    1.0  97.3
*13       +X                                                 5   507    1.0  98.3
*14       +X                                                 3   510    0.6  98.8
*15       +                                                  2   512    0.4  99.2
```

(d). Histogram of all new word waiting times of Martin Luther King's speech.

Figure 1 continued. **Note:** All histograms are plotted on the *same scale*. Frequencies of less than 5 do not appear.

## DATA

The data for this paper are drawn from a number of works by noted authors and speakers (Table 1). In all, twelve samples were selected, seven of which are noted oratory or prose, and five of which are monologues. The first group (Items 1 to 7 in Table 1) begins with two samples of oratory—Martin Luther King and Patrick Henry—followed by five of prose. As for the second group, the first four monologues (Items 8 through 11) were given by the male and female members of couples studied elsewhere for the properties of emotional dialogues. The final monologue (Item 12) is a free narrative given by a subject in response to a 30 second advertisement. The data selected is thus quite diverse.

Table 1. Authors and speakers.

| Sample | Content |
|---|---|
| 1. | King, I had a dream speech |
| 2. | Henry, Give Me Liberty or Give Me Death |
| 3. | Bacon, Of Friendship |
| 4. | Addison, The Vision of Mirza 1711 |
| 5. | Defoe, The Education of Women |
| 6. | Swift, A Treatise on Good Manners and Good Breeding |
| 7. | Emerson, Gifts (1844) |
| 8. | Monologue 1 |
| 9. | Monologue 2 |
| 10. | Monologue 3 |
| 11. | Monologue 4 |
| 12. | Monologue 5 |

In each case, the word sequence is converted to an integer sequence by using word position as follows: the first word and all of its occurrences are replaced by the number 1, the second word—if different from the first—and all of its occurrences is replaced by the number 2, and so on.

For example consider the first few lines from The Soliloquy of Hamlet (Act iii, Scene 1 of The Tragedy of Hamlet):

To be, or not to be: that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take up arms against a sea of troubles,
And by opposing end them. To die: to sleep;.

The text is replaced by the integer sequence:

1 2 3 4 1 2 5 6 7 8
9 10 11 12 7 13 1 14
7 15 16 17 18 19 20
3 1 21 22 23 24 25 26 18 27
16 28 29 30 31 1 32 1 33.

The above procedure transforms the original narrative from a sequence of words to a sequence of positive integers, with initial value 1. The distances between the occurrences of new integers in the sequence contain information on the generation of new words.

## METHOD: DATA ANALYSIS

Let $t$ be the number of words in a sample of text to a given point and let $h_C(t)$ denote the histogram of waiting times between new words up to time $t$. The subscript $C$ is used to suggest cumulative because the estimate of an exponential fit at time $t$ is based upon a histogram that has presumably accumulated information from prior exponentials. For each $t$, we compute a best exponential fit to $h_C(t)$ (see footnote[1]) and denote it $\lambda_C(t)$. The computed value of $\lambda_C(t)$ is the one giving the greatest confidence for goodness of fit using the Kolmogorov-Smirnoff (KS) criterion [6–8]. $\lambda_C(t)$ is well fit, via nonlinear regression[2], by functions of the form: $\lambda_C(t) = \lambda_1 e^{-\lambda_2 t} + \lambda_3 e^{-\lambda_4 t} + \lambda_5$. The true rate function $\lambda(t)$ is inferable from $\lambda_C(t)$. Figure 2 plots the regression of $\lambda_C(t)$ for the King speech and is typical of all others in our sample.

We take the present process as a discrete approximation to a continuous one and assume that $\lambda(t)$ is a continuous function defined on a closed interval. Suppose a continuous process is observed at discrete times $kT/N$, where $k = 1, 2, \ldots N$ and $T$ is fixed. At time $kT/N$ the expected quantity to be added to $h_C(t)$ is $\lambda(kT/N)/N$. Multiplying by $T/T$, summing over $k$ and letting $N$ approach infinity leads to:

$$\lambda_C(T) = \frac{\int_0^T \lambda(t)\,dt}{T}, \quad \text{or } \lambda(T) = (T\lambda_C(T))'.$$

A more careful analysis observes that the quantity added to $h_C(t)$, at time $kt/N$, is $b_k/N$, where $b_k$ is a random variable, that the limit processes of summation ($N$) and expectation ($b_k$) are presumed to commute, and that an ergodicity assumption is needed to assure the equality of the commuted limits. While we have no direct evidence for ergodicity, we note that human communicative processes are deeply stable and that the unfolding of any one realization tends to avoid abrupt or otherwise swift change.

The integral form of the above equation implies, as one hopes, that $\lambda(t) - \lambda_C(t)$ approaches 0 as $t$ approaches 0. The differential form implies that if $\lambda_C(t)$ is constant—as is the case for an ordinary Poisson process—then so is $\lambda(t)$. Thus, the above equations do satisfy some of the first necessary conditions that come to mind in order for them to be correct.

Figures 3A and 3B graph both $\lambda_C(t)$ and $\lambda(t)$ for the King data and for Monologue 3. At $t = 0$, $\lambda_C(t)$ and $\lambda(t)$ are equal, as they should be. As $t$ grows, $\lambda_C(t)$ falls below $\lambda(t)$, and for

[1] Each $h_C(t)$ had a correlation of .98 or better with the histogram data used to compute it.
[2] Nonlinear computations were done with the BMDP statistical package.

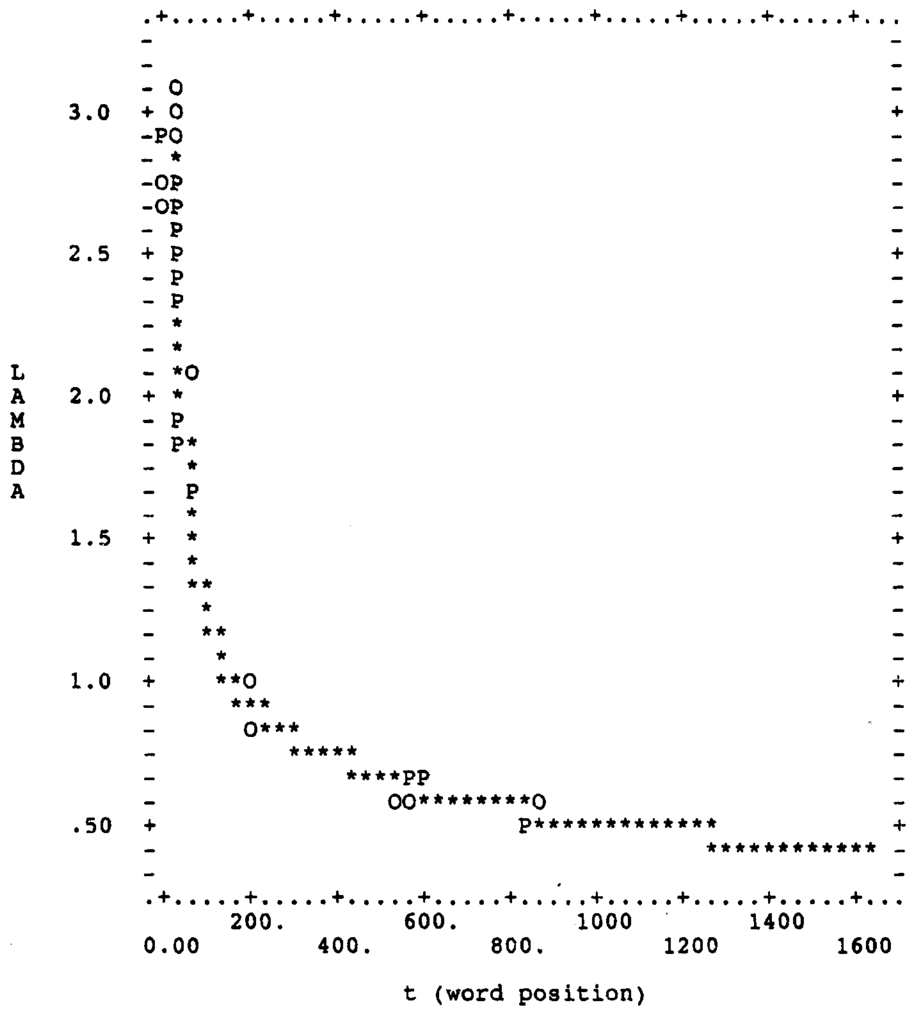



Figure 2. $\lambda_C(t)$ for Martin Luther King's speech. $P$ denotes predicted by the regression. $O$ denotes observed. $*$ denotes agreement of $P$ and $O$ to within plot resolution.

large $t$, $\lambda(t)$ approaches from below the far field behavior of $\lambda_C(t)$—asymptotic to $\lambda_5$. Given that $\lambda_C(t)$ is, in a sense, an acceleration of $\lambda(t)$, this geometry is to be expected.

## RESULTS

Table 2 reveals that the cumulative rate constant $\lambda_C(t)$ is well fit by

$$\lambda_C(t) = \lambda_1 e^{-\lambda_2 t} + \lambda_3 e^{-\lambda_4 t} + \lambda_5, \qquad \text{where } \lambda i > 0, \quad i = 1, \ldots, 5.$$

Indeed, the RMS's are uniformly less than 1/3% of the asymptotes. Using the differential equation above this implies that the true rate function $\lambda(t)$ is:

$$\lambda(t) = \lambda_1(1 - \lambda_2 t)e^{-\lambda_2 t} + \lambda_3(1 - \lambda_4 t)e^{-\lambda_4 t} + \lambda_5.$$

The derivative of $\lambda(t)$ is:

$$\lambda'(t) = \lambda_1\lambda_2(-2 + \lambda_2 t)e^{-\lambda_2 t} + \lambda_3\lambda_4(-2 + \lambda_4 t)e^{-\lambda_4 t}.$$

Table 2. Nonlinear regressions on raw estimated rate constant.

| Work | Word Length | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | RMS | DF | R | $N_z$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 104 K | 1641 | 2.6382 | −0.0244 | 0.7288 | −0.0020 | 0.3978 | 0.00168 | 1622 | .9928 | 3 |
| 108 H | 1220 | 1.5541 | −0.0098 | 0.7192 | −0.0004 | 0.0287 | 0.00133 | 1199 | .9931 | 3 |
| 107 B | 2590 | 2.7878 | −0.0293 | 0.5557 | −0.0013 | 0.3658 | 0.00208 | 2576 | .9853 | 3 |
| 109 A | 1795 | 1.8230 | −0.0267 | 0.6973 | −0.0021 | 0.3843 | 0.00119 | 1783 | .9928 | 3 |
| 110 D | 1406 | 1.5481 | −0.0348 | 1.0077 | −0.0034 | 0.3770 | 0.00087 | 1392 | .9958 | 1 |
| 111 S | 1945 | 1.9742 | −0.0387 | 0.6724 | −0.0023 | 0.4590 | 0.00102 | 1922 | .9887 | 3 |
| 112 E | 1509 | 1.9958 | −0.0236 | 0.7133 | −0.0017 | 0.3899 | 0.00138 | 1485 | .9915 | 3 |
| 044 M | 1756 | 1.4318 | −0.0183 | 0.9129 | −0.0027 | 0.3108 | 0.00092 | 1743 | .9961 | 1 |
| 045 M | 3511 | 6.7154 | −0.0498 | 0.7820 | −0.0014 | 0.3076 | 0.00118 | 3481 | .9900 | 3 |
| 046 M | 1870 | 1.0711 | −0.0146 | 0.6121 | −0.0022 | 0.3099 | 0.00076 | 1857 | .9942 | 1 |
| 047 M | 1720 | 2.8870 | −0.0531 | 0.8972 | −0.0026 | 0.3800 | 0.00074 | 1698 | .9560 | 3 |
| 083 M | 995 | 3.6885 | −0.0911 | 0.6749 | −0.0026 | 0.2626 | 0.00078 | 981 | .9930 | 3 |

DEFINITIONS.

**Word Length:** the number of words in the work. $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$: parameters of the regression model:

$$\lambda_C(t) = \lambda_1 e^{-\lambda_2 t} + \lambda_3 e^{-\lambda_4 t} + \lambda_5.$$

**RMS:** residual mean square of the regression.
**DF:** degrees of freedom.
**R:** correlation between fitted and observed data.
$N_z$**:** the number of zeros of $\lambda'(t)$ where

$$\lambda(t) = \lambda_1(1-\lambda_2 t)e^{-\lambda_2 t} + \lambda_3(1-\lambda_4 t)e^{-\lambda_4 t} + \lambda_5, \text{ and}$$
$$\lambda'(t) = \lambda_1\lambda_2(-2+\lambda_2 t)e^{-\lambda_2 t} + \lambda_3\lambda_4(-2+\lambda_4 t)\,e^{-\lambda_4 t}.$$

Thus $\lambda(t)$ is decreasing when $t < 2\min(1/\lambda_2, 1/\lambda_4)$, increasing when $t > 2\max(1/\lambda_2, 1/\lambda_4)$ (Figures 3A and 3B) and $\lambda'(t)$ can vanish only in $(2/\lambda_2, 2/\lambda_4)$, where we adopt the convention that $\lambda_2 > \lambda_4$. In particular, $\lambda'(t)$ is 0 when:

$$\left(\frac{\lambda_1\lambda_2}{\lambda_3\lambda_4}\right) e^{(\lambda_4-\lambda_2)t} = -\frac{2-\lambda_4 t}{2-\lambda_2 t} \quad \text{or}$$
$$\ln\left(\frac{\lambda_1\lambda_2}{\lambda_3\lambda_4}\right) + (\lambda_4-\lambda_2)t = \ln\left(-\frac{2-\lambda_4 t}{2-\lambda_2 t}\right).$$

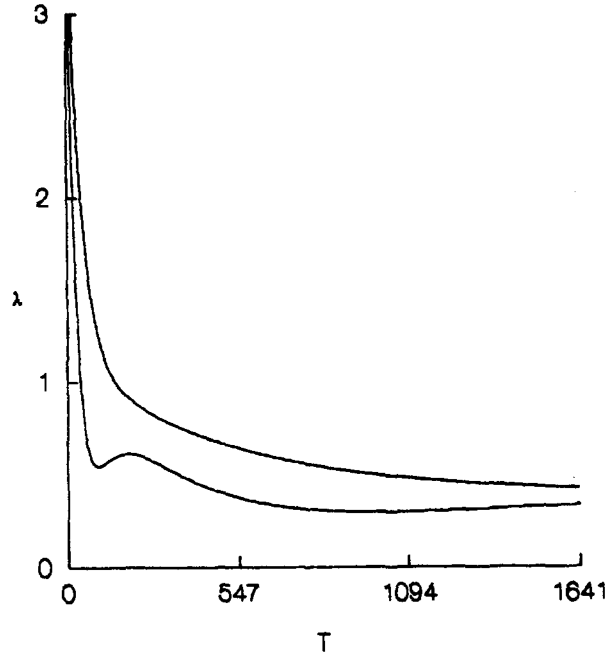
The zeros of $\lambda'(t)$ can be estimated by graphing the linear and ln function above together. Such a graph makes it clear that $\lambda'(t)$ can vanish 1 or 3 times (Figure 4).

In order for $\lambda'(t)$ to vanish twice the linear function must have the unlikely geometry of being tangent to the ln function at some point. This is an interesting degenerate case because if $\lambda'(t)$ vanishes one or three times, then $\lambda(t)$ approaches its asymptote—$\lambda_5$—from below, but if $\lambda'(t)$ vanishes twice then $\lambda(t)$ approaches its asymptote from above. In the present data, *all derivatives vanished either one or three times*, a point we discuss below.

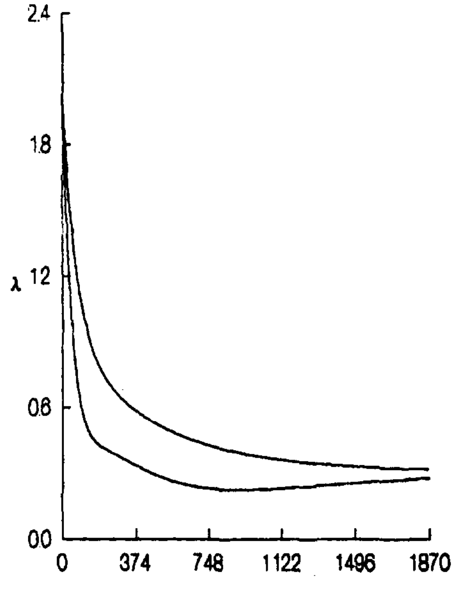
When $\lambda'(t)$ vanishes three times the geometry of $\lambda(t)$ is complex, involving two local minima and one local maximum, a notch configuration as in Figure 3A. On the other hand, when $\lambda'(t)$ vanishes once the geometry of $\lambda(t)$ is simpler and more graceful as in Figure 3B.

## PHASES OF NEW WORD GENERATION

The form of $\lambda(t)$ implies that the generation of new words, written or spoken, has a beginning, middle and final or asymptotic phase. Figure 3 suggests an almost piecewise linear model for the phases. All three are operative at $t = 0$, with the first and then the second becoming essentially extinct after some time.



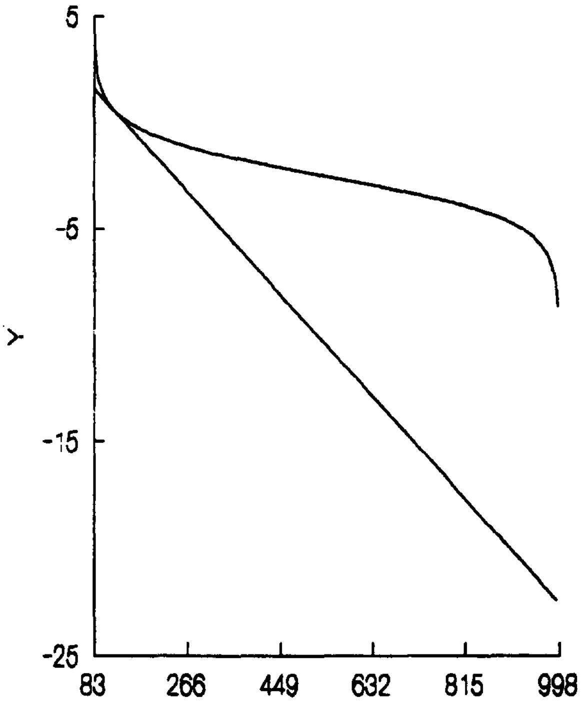

(a). The cumulative rate, $\lambda_C(t)$ and the true rate, $\lambda(t)$ for Martin Luther King's speech. $\lambda_C(t)$ is the upper plot.



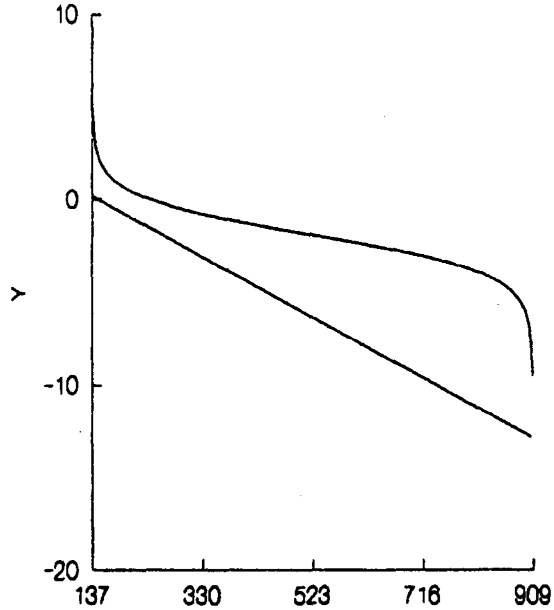

(b). The cumulative rate, $\lambda_C(t)$ and the true rate, $\lambda(t)$ for Monologue 3. $\lambda_C(t)$ is the upper plot.

Figure 3.

There are no instances in the present data where $\lambda'(t)$ vanishes twice. As noted, this corresponds to approaching the asymptote from above, an event of ongoing fatigue. When $\lambda'(t)$ vanishes 1 or 3 times, the far field behavior of $\lambda(t)$ is to approach its asymptote, $\lambda_5$, from below. There is, thus, something of an overall U-shape to $\lambda(t)$. This is a familiar phenomenon in human task activity and is usually interpreted to signify entry into an activity with gusto, followed by fatigue, and concluding with fresh energy in anticipation of quitting the task.




(a). Plot of $\ln(\lambda_1\lambda_2/\lambda_3\lambda_4) + (\lambda_4 - \lambda_2)t$ and $\ln(-(2 - \lambda_4 t)/(2 - \lambda_2 t))$ to locate the number of zeros of $\lambda'(t)$ for Martin Luther King's speech. **Note:** There are 3 points of intersection. Two are in the upper left hand corner, and one in the lower right hand corner where the linear function meets the (unplotted) asymptote to $-\infty$ of the ln function.




(b). Plot of $\ln(\lambda_1\lambda_2/\lambda_3\lambda_4) + (\lambda_4 - \lambda_2)t$ and $\ln(-(2 - \lambda_4 t)/(2 - \lambda_2 t))$ to locate the number of zeros of $\lambda'(t)$ for Monologue 3. **Note:** There is 1 point of intersection in the lower right hand corner where the linear function meets the (unplotted) asymptote to $-\infty$ of the ln function.

Figure 4.

When $\lambda'(t)$ vanishes 3 times—the notch in $\lambda(t)$ as in Figure 3A—there is a "local" U at the transition from phase one to two. This corresponds to exit from phase one, a fatigue event, followed by recovery with entry into phase two. Overall then, one may wish to infer that both phase transitions—one to two and two three—begin with fatigue and conclude with recovery into the next phase. In this case, there are two U's, one for each transition. These tentative inferences open the question as to what does a single zero of $\lambda'(t)$ signify.

## NUMERICAL RANGE

The parameters in Table 2 are tightly clustered and give little surface suggestion as to whether $\lambda'(t)$ vanishes 1, 2 or 3 times. The question must be advanced: is the presence or absence of the notch (3 zeros for $\lambda'(t)$) a result of the method of estimation? It is well known that small changes in the parameters for a sum of exponentials fit can produce large changes in the resulting fit.

We believe that the *absence* of the notch—a single zero for $\lambda'(t)$—is the result of ill conditioning. There is a uniform recovery from the middle phase to the end phase in every case. This is the longest regime in the data, commencing at about $2/\lambda_2$ as compared to the interval from $t = 0$ to $2/\lambda_2$ where the first transition can be identified. That is, it appears to be in the nature of the process to provide enough data to fully identify the second phase transition but not always the first. Prior experience indicated that at least 600 words are necessary to estimate $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, and $\lambda_5$. The present samples of text, ranging from 995 to 3511 words, favor more accurate estimation of the middle to final phase.

In addition, as illustrated in Figure 3B, where there is no notch, there is still a tumescence (an interval where $\lambda''(t) < 0$) in the graph of $\lambda(t)$—suggesting that the sensitivity of the estimation method, and/or an insufficiency in the amount of data, is obscuring a real notch.

The U-shaped performance curve is well known and is expectable in a fresh domain. We also believe for this nonmathematical reason that the absence of the notch implies an estimation problem versus a distinctive feature. Finally, we are convinced too that the unanimous absence of data for which $\lambda'(t)$ vanishes twice speaks for itself in arguing that $\lambda'(t)$ has 3 zeros.

## INDIVIDUAL DIFFERENCES

The present sample is too small to support the inference of significant differences. However, there are suggestions in Table 2 of differences between the noted authors and speakers (1–7) as compared to the anonymous monologues (8–12). With the exception of Patrick Henry there is some tendency for the first group to have higher asymptotes the second group. Table 2 also suggests that the second group has more variety in the parameters 1 and 2 than the second.

## CONCLUSION

The generation of new words in spoken or written text appears to be an evolving Poisson process having a distinct beginning, middle and final phase. The functional form of the time dependent rate $\lambda(t)$ reveals that the new word rate begins with a relatively large opening burst, after which it declines. The decline is sometimes monotonic and is otherwise interrupted by a brief notch in phase transition from beginning to middle.

The finding of an evolving Poisson process for new words is continuous with prior results. We found this process in studying monologues and the members of emotionally charged dialogues [9–11]. We argue that if another observable, contained within word sequences, defines a Poisson process then it is likely to do so as a macro-recapitulation of a more primitive Poisson process. Indeed, we conceptualize the new word Poisson process as a "carrier" or even a generator for higher order processes whose definitions are contained within the words themselves.

No cases were observed for which $\lambda'(t)$ vanishes twice. Thus, there is no data for which $(t)$ approaches its asymptote from above. We take this as evidence that humans generate new words without asymptotic fatigue and note that this also characterizes most known human tasks.

The new word generalized Poisson process is nonstationary but is probably independent of its initial point in a long sample of text. More data and analysis are required to verify this conjecture. This form of translation invariance, if verified, would suggest that the process has some ergodic properties.

## REFERENCES

1. A. Badalamenti, R. Langs and J. Robinson, *Behavioral Science*, Volume 39, pp. 46–71, (1993).
2. D. Cox and H. Miller, *The Theory of Stochastic Processes*, John Wiley, New York, (1965).
3. A.I. Khinchin, *Mathematical Methods in the Theory of Queuing*, Griffin & Company, London, (1969).
4. T.C. Saaty, *Elements of Queuing Theory*, McGraw-Hill, New York, (1961).
5. A. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Dover, New York, (1973).
6. M. Fisz, *Probability and Mathematical Statistics*, John Wiley, New York, (1967).
7. M. Hollander and D. Wolfe, *Nonparametric Statistical Methods*, John Wiley, New York, (1973).
8. M. Kendall and Stuart, *The Advanced Theory of Statistics, Volumes I-III*, Hafner, New York, (1968).
9. A. Badalamenti and R. Langs, An empirical investigation of human dyadic systems in the time and frequency domains, *Behavioral Science* **36** (2), 100–114, (1990a).
10. A. Badalamenti, R. Langs and M. Kessler, Stochastic progression of new states in psychotherapy, In *Statistics in Medicine, Vol. 11*, pp. 231–242, (1992).
11. R.J. Langs, A.F. Badalamenti, Stochastic analysis of the duration of the speaker role in psychotherapy, *Perceptual and Motor Skills* **70**, 675–689, (1990a).