READING TO THE ENDGAME
D. GRAHAM BURNETT & W. J. WALTER

THE PROBLEM

In his brief essay "Gli scacchisti irritabili" ("The Irritable Chess Players") of 1985, Primo Levi elaborates a set of symmetries between the act of literary creation and the playing of a game of chess. Both a work of literature and the royal game, he suggests, unfold in time within strictures that inexorably invoke "life and the struggle for life." There is, as he puts it, a "symbolic shadow" that lengthens over a chess board, since the way to the end is the way to a death, "a death for which you yourself are guilty." The novel, of course, is the literary form that has evolved precisely to afford language the means of erecting and choreographing such a metaphorical life space. And thus it is no surprise that the novel, too, is haunted by a long shadow: all plots, as Don DeLillo memorably put it, end in death. Moreover, en route to their respective endgames, both chess and the novel offer powerful arenas in which to investigate the question of questions: the ever-vexatious issue of the relationship between fate and agency, between necessity and freedom. Every move is our own, except when it's not. Either way, the board thins, the sheaf of paper in the right hand dwindles, sifting left as if blown by an inexorable wind—though of course, we turn every page. Chess, in this sense, is the opposite of dice, just as the novel is the opposite of Scripture (the exact difference between chance and providence has never been clear, but they share an antithesis in deliberative subjectivity, and this may be a clue).

Stimulated by Levi's juxtaposition, and motivated by the possibilities of extending an Oulipian sensibility into the sphere of literary criticism (OuCriPo?), the authors set out to develop a means by which a given novel could express itself as a game of chess. Initial success here led to expanded ambition, since there was nothing to stop us from elaborating our modest analytic protocol into a full-fledged "engine" that would permit works of literature to confront one another on the chess board. We have advanced this project to what we think of as a workable tool for a certain sort of ludic literary investigation, and we present it here for the first time, together with some preliminary results drawn from several thousand games we have run to date. The current version of the program is playable on the *Cabinet* website, and we would be delighted if it proved useful to those wishing to pursue this or related lines of inquiry.

FROM CHARACTERS TO MOVES

A chessboard consists of sixty-four squares commonly designated by alphanumeric coordinates (a-h across the x-axis and 1-8 up the y-axis). If one were to replace the numerical assignations with a continuation of the alphabet (running, for instance, i-p up the y-axis), each square would be designated uniquely by a two-letter coordinate that we will call a "tuple." Now imagine setting up a simple computer program that knows the rules of chess—nothing more. It knows, for instance, all the moves that are makeable by a given piece, and it can keep track of a chessboard (updating what pieces are on which squares as moves are made). Suppose further that this program takes directions for making moves in the form of a pair of "tuples"—namely, one letter-pair designating the coordinates of a square occupied by a movable piece, and then a second letter-pair designating the coordinates of a square to which that piece can be legitimately moved (including squares where it would capture an opposing piece).

We now have everything in place to convert two texts into a game of chess: we simply feed the program the two novels, asking it to play one text as "white" and the other as "black"; the program searches through the white text until it finds the first tuple corresponding to a movable piece (in the case of an opening move, either a pawn or a knight), and then, having settled on the piece that will open, continues searching through the text until it encounters a tuple designating a square to which that piece can be moved. When it has done so, the computer executes that move for white, and then goes to the other text to find, in the same way, an opening move for black. And so it goes: white, black, white, black, until—quite by accident, of course, since we must suppose that the novels know nothing of chess *strategy* (and our program cannot help them, since it knows only the *rules* of the game)—one king is mated.

Such a set up would be close (there turn out to be interesting differences, but put that aside for now) to permitting two monkeys to play chess against each other by giving each a keyboard and permitting them to jump about on them: send the resulting string of letters to our program, and it scans this string of gobbledygook for tuples that constitute legitimate moves, makes them, and *voilà*, monkey chess.

We experimented with something along these lines (no monkeys, as it happens, but a similar sort of stochasticity) before setting to work building into the basic application a greater sensitivity to the specific qualities of the string of characters upon which it was set to work. Our aim was to produce an algorithm that could, in

some modest way, "read" literary works for their distinctive "voice" or "style" and then convert that linguistic particularity into something approaching a style of play.

STYLOMETRICS

In this undertaking, we took our inspiration from the bastard semi-discipline known as "stylometrics." While assigning numerical values to letters and words is an ancient practice (the Greeks even had a word for this activity, *isopsephy*, which they linked to divination), the use of quantitative methods to characterize stylistic features of a given work or author is a surprisingly recent technique of literary analysis. There were some quirky feints in this direction by German biblical scholars in the early nineteenth century (including Schleiermacher, who counted up the instances of exceptional phrases in one of the Epistles by way of calling into question Pauline authorship). Shortly thereafter, classical philologists also began to mobilize simple statistical arguments about word usage as part of various arguments about the chronology and authenticity of canonical works by Plato and others. It was not until the 1970s, however, that the widening availability of programmable computers opened the way to computationally intensive forms of stylistic analysis, and by the mid-1980s these techniques had achieved modest notoriety as a result of apparent victories in several public controversies (in the United States, much attention focused on determining authorship of *The Federalist Papers*, which has become a kind of lab-rat for new stylometric techniques; in Europe, Mikhail Sholokov's *And Quiet Flows the Don* was prominently sliced and diced by a brace of mathematically sophisticated Russian scholars with baroque political axes to grind).[1]

It is the premise of stylometry that writers necessarily, albeit unintentionally, imprint their texts with distinctive statistical signatures, which can be identified by means of close attention to sentence length, word and letter frequency, and other quantitative attributes of their prose. Thus a chess program sensitive to the stylometric parameters of a given text should in principle afford a mechanism for translating authorial voice into a way of playing the game.

THE ALGORITHM

Our working model of such a system is, we freely admit, somewhat primitive and more than a little arbitrary, especially given the imaginable possibilities. But we find it suggestive nevertheless. Let us say that we wish to pit Jules Verne's pre-Freudian chthonic womb-fantasy *Voyage au Centre de la Terre* against Jane Austen's

paradigmatic quasi-romance *Sense and Sensibility*. Presented with these texts, our program sets to work on a basic stylometric analysis of the opponents, compiling a list of the sixty-four most commonly used tuples in each text by rank of their frequency. The machine then assigns these tuples, from most common to least, to the sixty-four squares of the board, proceeding in a spiral from the four central positions.

White Spiral

	a	b	c	d	e	f	g	h
8	50	51	52	53	54	55	56	57
7	49	26	27	28	29	30	31	58
6	48	25	10	11	12	13	32	59
5	47	24	9	2	3	14	33	60
4	46	23	8	1	4	15	34	61
3	45	22	7	6	5	16	35	62
2	44	21	20	19	18	17	36	63
1	43	42	41	40	39	38	37	64

Black Spiral

	a	b	c	d	e	f	g	h
8	64	37	38	39	40	41	42	43
7	63	36	17	18	19	20	21	44
6	62	35	16	5	6	7	22	45
5	61	34	15	4	1	8	23	46
4	60	33	14	3	2	9	24	47
3	59	32	13	12	11	10	25	48
2	58	31	30	29	28	27	26	49
1	57	56	55	54	53	52	51	50

Assigning Tuples by Frequency: The sixty-four most frequent letter pairs are assigned to the squares of a chessboard starting from 1 (most frequent) to 64 (least).

The tuples of highest frequency (in Verne's novel, *es*, *en*, *le*, *on*; in the Austen, *he*, *th*, *er*, *in*) thus control the strategic center of the board. Instead of the old, arbitrary tuple coordinates based on a-h and i-p, the squares on the board are now "called" by tuples that have a patterned and systematic relationship with the work in question. And note that these positions are assigned separately for each novel, so each text will play on a coordinate field determined by its own linguistic parameters.

We are now ready to begin play, with the program seeking an opening move for white by scanning, as already outlined, first for a tuple corresponding to a movable piece, and then, having made that determination, for a subsequent tuple corresponding to a legitimate move. In this case, white finds the unconventional-but-storied "Grob attack" or "Spike opening" in its very title (after a heady debate we decided to treat the title and author, though not the publisher or place of publication, as "part" of the analyzable text): the "au" of *Voyage au Centre* having been assigned in this case to g2, and thus to white's g-pawn; and the "tr" of *Centre* designating, again on the basis of the initial tuple frequency analysis, the board position g4. Black, seemingly wobbled by this irregular attack, answers with a hesitant Rook-pawn to h6, on the basis of the "li" in *Sensibility* and the "fo" in the telling opening phrase "for many generations." The contest is joined.

It plays out as depicted below, with white uncorking a fatal King-pawn to d4 on the basis of the palindromic tuples "se" and "es" as they appear in the following sentence:

Mais de faire entendre raison au plus irascible des professeurs, c'est ce que mon caractère un peu indécis ne me permettait pas. (But to make the most irascible of professors hear reason, this my somewhat hesitant character did not allow.)

|	White	Black
1.	g2–g4	h7–h6
2.	e2–e4	f7–f6
3.	e4–e5	e7–e6
4.	e5xf6	g7–g5
5.	g1–f3	e8–f7
6.	d1–e2	b8–a6
7.	f1–h3	a8–b8
8.	f3–e5+	f7xf6
9.	c2–c4	h8–h7
10.	h1–f1	h6–h5
11.	b1–c3	c7–c6
12.	f1–g1	f6–g7
13.	e2–e4	b8–a8
14.	c4–c5	b7–b5
15.	c3–e2	g7–f6
16.	e2–c3	b5–b4
17.	a2–a4	b4xc3
18.	g1–g2	c8–b7
19.	e4–f3+	f6xe5
20.	e1–e2	d7–d6
21.	e2–d3	a6xc5+
22.	d3xc3	h5–h4
23.	c3–c4	c5–e4
24.	g2–g1	g8–f8
25.	d2–d4#	
	1–0	
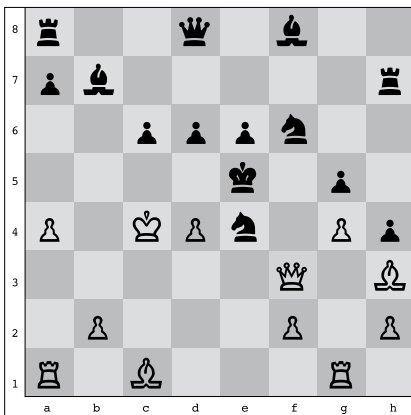
***Voyage au Centre de La Terre* (white) traps *Sense and Sensibility* after two careless Knight moves hem in black King at e5.**

PRELIMINARY RESULTS AND DISCUSSION

Over the last several months we have experimented with fifty-five works in five languages (English, German, French, Italian, and Spanish), representing seven

national literary traditions (distinguishing American and English writers, and Spaniards from Mexicans), and tending to emphasize classic novels of the nineteenth century. (We are currently working on a version of the program that can handle Cyrillic characters, since we have felt the absence of Russian sources, particularly when running international tournaments). Using batch processing and many machine-hours, we have logged complete records of several thousand games, and have analyzed this data set a number of different ways. Some general observations will give the reader a feel for these results.

For starters, a relatively brief, "tight" game like *Voyage au Centre de la Terre* vs. *Sense and Sensibility* is more the exception than the rule: overall about 30% of our games end in checkmate, and the rest wind up as draws (of which a little less than 30% are stalemates).[2] Of the games that end with a victor, the average number of moves to mate is seventy-one, and in many of these pawn-promotion plays a significant role (pawns reaching the opponent's home row can be "promoted" to Queen, and several "Queens" of the same color can thus occupy the board at the same time).[3] Our longest contest to date is the marathon showdown between Herman Melville's tricky *The Confidence-Man* (1857) and the novel often accorded foundational status in the history of *el realismo* in Spain, *La Gaviota* (*The Seagull*, 1849), authored by the cosmopolitan aristocrat Cecelia Böhl von Faber (writing under the pseudonym Fernán Caballero). This *tête-à-tête* sees von Faber's celebrated *naturalismo* grind Melville's Manichean insouciance to the humiliating configuration shown here after no fewer than 159 moves.
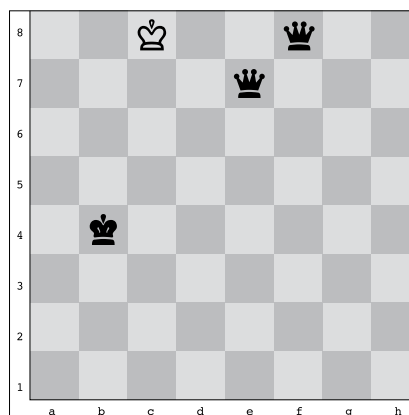
***The Confidence-Man* finally cornered by the picaresque *La Gaviota* after a lengthy chase, which concluded black Queen e7-e6+; white King a6-b7; black's other Queen c2-f2, b7-b8; e6-e7, b8-c8; followed by the deep strike f2-f8#.**

Nearly as exhausting is the 141-move slugfest between Benjamin Constant's lapidary tale of ennui and seduction, *Adolphe* (1816), and *La Bodega* (1905), Vincente Blasco Ibáñez's crusading proto-prohibitionist indictment of wine-drenched backwardness in Andalucía—a game that culminates in yet another triumph for Iberian social realism. These contests stand in marked contrast to the Teutonic lightning-strike by which Theodor Fontane's *Effi Briest* (1894)—a favorite of Thomas Mann's—undoes Rafael Delgado's slightly over-ripe late exercise in Mexican sentimentality, *Angelina* (1947): an eleven-move frontal *Blitzkrieg* that ends as follows:
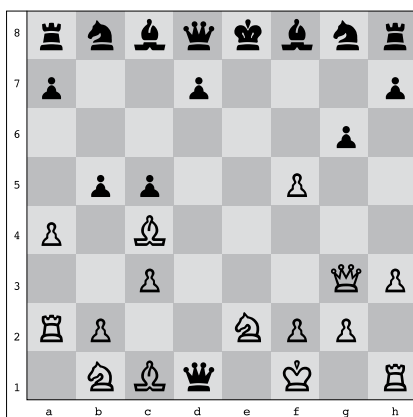
A variation of the "Prussian Pawn": *Effi Briest* (playing black) goes straight up the gut against an inattentive *Angelina*, the crucial move being black King-pawn e3xd2 at move eight, on the basis of "*grün quadrierten*" and "*und dann über diesen hinaus.*"

We were prepared to believe, reviewing the transcripts of many considerably more meandering games, that what we were looking at was really—in the end, putting aside the vague charm of thinking of the opponents as works of literature—indistinguishable from the sort of aleatory "monkey chess" alluded to above. But a control seems to prove otherwise: we built a small test module of our program that did indeed choose moves for black and white on a random basis, and discovered to our surprise that this device posted a significantly *higher* rate of checkmate games than actual novel chess (slightly over 40%). We have an intuition as to why this should be so, but it is tentative, and we reserve it at present, welcoming in the meantime suggestions from those interested in taking the question up.

Other peculiarities remain. In something akin to match play, we find French novels in general, and *Adolphe* in particular, to be the toughest competition. This would doubtless delight François Le Lionnais, the

chemical engineer who initiated the Oulipo group and authored half a dozen books of chess problems, but we are at present without an explanation as to why it should be so. In our last "world cup," a five-novel French team (led by Alexandre Dumas) trounced the competition, defeating the second-place English team by a 20% margin (the UK was hurt by poor performance from Disraeli's *The Infernal Marriage* of 1834), and leaving the also-ran Italians looking rather like novices. Sample sizes for this tournament were not enormous (on the order of 900 games, resulting in some 300 checkmates), and we are continuing to explore just how statistically robust this apparent anomaly actually is. We are inclined to think that differences in "national styles"—if firmly established—will be explained by reference to the striking, and seemingly consistent, divergence of tuple frequencies across different language groups, divergences particularly pronounced at the high end of the frequency range. Grossly speaking, the four most common tuples in a given French text tend to be relatively close in their overall rate of appearance. By contrast German, for instance, sees a sharper decrease in frequency across this same range (interestingly, English, which on average falls somewhere in-between, shows a much higher degree of *variability* in this metric). How exactly these sorts of statistical characteristics affect game performance under our algorithm we cannot yet say.

As to why Constant's thinly veiled exposé of the intricate erotic politics surrounding Madame de Staël should prove so formidable an adversary, that is a perfect mystery (though it should be noted that both *Frankenstein*, in the second edition of 1831, and Goethe's epochal 1774 *Die Leiden des Jungen Werther*, playing black, successfully fight off forceful opening gambits and record victories against this opponent). For those who remember without fondness the distinctively suffocating adolescent solipsism of *Adolphe*, there is a perverse satisfaction to be had watching *Adolphe* (playing black) fall to *Adolphe* (playing white) in fifty-nine shambolic moves that involve much fruitless horseplay. An appetite for such forms of poetic justice is not, however, consistently sated by our software, as one certainly does not expect *Huckleberry Finn* (1884, playing white) to be picked apart in passive resignation by Kafka's *Die Verwandlung* (*The Metamorphosis*, 1915), and it is difficult to look on with suitably clinical detachment. In particular, a too-little, too-late counterpunch of (promoted) Queen b8-e5 must be said to painfully darken the miasma of futility that hangs over this confrontation from the first move (Huck's defensive "Anderssen" opening, the paralytic a2-a3).

In closing, it is perhaps worth addressing the possibility of *composing* novels specifically calibrated to win chess tournaments convened by our application. The prospect of a "grand-master," a kind of all-purpose novel-killing novel, while alluring, strikes us as beyond reach—indeed as probably a formal impossibility. But a novel written to defeat some other specific novel would appear to be an attainable objective, though a little thought suggests this would be by no means a trivial undertaking. While "encoding" tuples that would defeat a given opponent would in principle be quite easy, the real problem is doing so in the context of a work that must generate, on the basis of stylometric analysis of overall tuple frequencies, the very coordinate plane upon which those moves will be made. This would seem to give the problem a recursive character of some complexity.[4] More work is needed.

1 For an overview of the state of the field in this formative period, see Anthony Kenny, *The Computation of Style* (Oxford: Pergamon Press, 1982). For a brief history of the development of the analytic techniques themselves, consider: David I. Holmes, "The Evolution of Stylometry in Humanities Scholarship," *Literary and Linguistic Computing*, vol. 13, no. 3 (1998), pp. 111–117.
2 Our program treats both three repetitions of the same board position and fifty moves without capture or pawn move as "automatic" draws; this is a slight tweak of chess rules, which require in each case that the draw be formally claimed by a player.
3 Serious chess players will want it noted that a pawn reaching the opponent's home row *may* be promoted to any board piece (except a King) and *may not* remain a pawn. In practice, our program promotes only to Queen. In human play, promotions to other pieces are very rare.
4 A saving grace here may lie in the fact that many of our games play out without making use of more than a small fraction of one or both of the novels at issue. This would seem to open the way to drafting a novel, analyzing its tuple frequency, building the board coordinates (using the spiral configuration outlined above), and then "editing" the opening chapter to plot the moves necessary to defeat a given opponent, *but doing so while maintaining tuple-frequency neutrality* (probably by off-sets elsewhere in the text).