

WRITING ON DISK

An Author at Play in a Computer-Simulated World

Jane Dorner

9 Collingwood Avenue, London N10 3EH

Synopsis: This paper suggests that computer manipulation of written material is insubstantial. Despite the huge choice of electronic writing aids, few match up to what is required by professional writers, editors or translators. The creation of texts on screens further desubstantiates the written word and forces copyright owners to redefine the 'copy' as a unit of sale. One of the many challenges facing those involved in publishing is to make provision for the implications of presenting and displaying work in electronic form.

1. Introduction

If there is such a thing as virtual reality, which is a computer-driven picture of the real world — apparently similar, but lacking the substance of the real thing — then, I would like to argue, there is virtual writing.

Real writing is what a person can understand, using anything from intuition to prejudice to sift out meanings from ambiguities. A person will almost always interpret, however garbled the syntax and grammar; however difficult the substance. *Virtual* writing is what a computer pretends to understand. It is a computer picture of the written language using algorithms and pattern-matching to find, display or analyse strings of text. This offers a play area in which the computer is being 'trained' in three broad directions that are of interest to the writer:

1. to offer shells for massaging creative ideas into literary forms (random plot generators, poetry processors, crossword puzzle makers);
2. to supply the need for immediate interactive reference (pop-up dictionaries, translation thesauri, CD-Rom encyclopedias);
3. to act as an interactive editor or translator by apparently making judgements based on internal program algorithms (grammar-checkers, style analysers, automatic translation programs).

In the research for my recently published book [Dorner (1)], I have played with 50-odd computer toys, some briefly and others in considerable depth. Many of these offer themselves as serious computer tools for professionals, with price tags to match (I estimate that the total value of software I have had for review is £5,000, exclusive of word-processors). Others have an educational value but are very constricted by memory limitations. Producers obviously see

this as a wide open market. The last two years has seen an enormous growth of such products; some new, others upgrades of previous products. Two years ago, a survey showed that tools such as these were in use by between 1% and 9% of professional authors [Dorner (2)]. This growth must be aimed at cost cutting. Figures released last year show that in most offices a standard letter costs a company £3.50 a page and documents originating from senior heads of department cost £50 a page (paper, printing ink, overheads, writing, reading and editing). Computer tools aim to help cut costs by reducing time spent in research and quick reference and by making it easier to attend to style on screen rather than on paper.

2. An Electronic Workshop

Those involved in writing lag behind musicians and composers who have embraced the computer for its creative and labour-saving opportunities. Our equivalent would be a multi-screen workshop where a writer could swivel round in a semi-circle to see five or six screens monitoring text at different planning, writing or editing stages — a visual synchrony of the mental processes that occur during writing. Imagine being able to turn to — or ignore — that screen which is checking your grammar, or showing the structure of your ideas, or providing keyword links to everything else you ever wrote on the same subject, or identifying changes made since the last draft, or simultaneously translating the work into five different languages. In this picture the writer does what the musician does with sound; knob-twiddles with words and ideas. In this way authors can play when inspiration lags, while accepting that final control always lies firmly in the operator's hands. It's just possible that too much dependence on the computer would be unhealthy, but how much that is to do with ergonomics and safe usage of these electrical toys is not under scrutiny here.

3. Rules and the Game

In virtual writing you simulate real writing just as in virtual reality you could simulate real tennis. The game offers a complete framework in which the text is bound by a large number of rules. These can be obeyed or broken or changed to fit a specific need. Such rules are computer-agitated — shaking the subject material until it fits the slots available, like ball-bearings rolling over a perforated board. There are no winners or losers; players are trying out their texts in this world of rules to see how they look or what can be learned from the process, much as an architect would rotate a computer-modelled house to explore uncharted corners before going back to the drawing-board.

The scoring in this game of virtuality is not numeric; it is the more subtle effect of the human player seeking to prove superiority over a digital foe. There is something about the way in which the computer reduces everything to an ephemeral series of pixellised graphemes, in which on and off light switches provide the only substance, that requires a special alertness on the part of the writer. One of the reasons for the success of the computer as a modern linguistic tool must be that human beings, in their arrogance, enjoy pitting their wits against it. The attraction, perhaps, of machine-assisted translation is not so much the time saved in situations where vocabulary is strictly limited — e.g. standard letters, business reports,

technical descriptions, business proposals, sales literature, bills of lading, instructions and directions — but the pleasure of post-editing and improving on the computer's best guesses. A simple example from the hard-disk-hungry program Globalink is the German 'In den letzten Jahren' which translates as 'in the last years' rather than 'in recent years'.

4. Fun Element

Scoring, of course, is not necessary in a simulation because the playing itself is a prop to learning. The modern trend in computer-driven writing makes such learning fun, with pretty pictures egging the user on to new discoveries about what each bit of software can do. People worry about the increase of this Noddy element on computer screens. The Delaware University research project — a project largely discredited by the academic community — that examined students' literacy level and choice of 'serious' subjects according to whether they used a Macintosh or an IBM running Dos apparently 'proved' that the friendlier icon environment fostered fluffier subject material and lower reading levels in the writing.

The worry is that, in Pope's phrase, 'Amusement is the happiness of those who cannot think' (and for the record, I found that epigram in my pop-up dictionary of quotations, sliding it into my text like jelly off a hot spoon). Pope, of course, is a moralist, but perhaps there is a case for more serious study on the effect of the amusement value of computer-aided writing on the individual than the Delaware project. My hunch would be that the debonair point-and-click flirtation ends with the print-out, when the reality of the marriage reasserts itself. It doesn't take long to learn the acronym GIGO — Garbage In, Garbage Out.

We need to decide if we can expect the computer to help writers, in the same way that virtual reality helps pilots to fly planes, or architects to design buildings. If it made a significant contribution of our understanding of the way language works, that would seem to be an achievement. However, it isn't clear to me, yet, that all the toys and gadgets I have looked at give us anything more than a very partial understanding of language. Sir Randolph Quirk, commenting on papers in a symposium on directions in corpus linguistics [Svartik (3)], reinforces a cautionary attitude. There is, he says:

a certain tension between (a) those who want to know as much as possible about language (and God bless the computer if it can help), and (b) those who want to know as much as possible about what the computer can do (and God bless it if this advances our knowledge about language).

Returning now to my three bands of computer help, can we 'bless the computer' at all for the help it offers? Names of products that I have played with are given in italics within each category.

5. Structural Help

Outliners, brainstorming and various kinds of conferencing software are in business usage. Some writers prefer to use an outliner to establish a structure and then flesh this out. There is

a fundamental assumption behind the concept of an outliner that is open to question. It is based on the claim that good prose is a consequence of planned rhetorical organisation: you do your thinking first and then plaster in the gaps. Set against that is the view to which I subscribe — good prose is a consequence of spontaneous discovery: you paint in the broad brushstrokes and find out what you want to say as the colour spreads. A colleague describes this as the 'romantic approach' as opposed to the 'classical approach' favoured by software grids [Galbraith (4)]. Controversial (and as yet unproven) as such a theory may be, it does at least have interesting implications for the design and use of software tools like outliners. Anything that offers a prompted system for helping the writer needs reviewing with some caution. It may be that outliners are useful for reorganising writing, but have serious limitations in the *generation* of ideas.

The kind of thing virtual writing requires is that a poem, a play or a genre novel writing can yield to algorithmic analysis: e.g. the story is the problem, the computer program offers the procedure for solving it. The simpler the medium, the more likely this is. Haiku can be effectively computer-generated. It is not obvious whether a person or a computer had written this:

All green in the leaves
I smell dark pools in the trees.
Crash, the moon has fled.

[Masterman (5)]

Random plot generation is another matter. Exploring a first thought that may have been randomly generated by typing answers to a questionnaire or clicking a mouse on a 'box' of themes is *one* way of thinking creatively. In creative writing classes (as in art and design) much is made of the value of developing random ideas and taking them beyond themselves. As an initial stirrer of the pot of ideas, a computerised story generator has as much to offer as any other technique, and very often more because it offers a playful exploration of the structures and processes of written language.

However, given the complexity of human fictions, handling them scientifically is bound to be extraordinarily difficult. I've played with a few of these things and am provisionally prepared to declare that as a writer, I have no interest in them. They seem to me mere toys offering stereotype solutions and unimaginative combinations. They probably have their place in education, because in playing 'what if' games a student learns to define 'what happens next'. On balance, though, I see no future for literature in structural tools.

Plots Unlimited, StorySpace, AIQ, Newman's Poetry Processor, Idealist, EndNote, Macrex, GrandView, Brainstorm, Thoughtline.

6. Pop-up Help

The second camp of linguistic interest is directed at converting paper books into efficient on-line reference. I do not mean multilingual databases such as the EC's Eurodicautom in Luxembourg, but more home-based products that replace reference books. The reduction of the 20 volumes of the *Oxford English Dictionary* 2nd edition or *Grand Robert Electronique* to

a thin laser-disk is a modern miracle. They cost less and are more useful than the paper equivalents because of the magical sophistication of the search routines. Less ambitious products comprising pop-up dictionaries of usage, quotations, reference books, bibliography lists, multilingual dictionaries, and a thesaurus or two illustrate a current growth area. Success here depends on software and hardware engineering rather than linguistic expertise. Most of these lurk in background memory and can seriously reduce the amount of text you may hold in active memory.

It is only a matter of time before computer memories are large enough and CD-ROM general enough and the pop-up complement to virtual writing will be commonplace. It will *feel* insubstantial to be checking a word-derivation or quotation or client via a series of light switches rather than hefting down a volume and thumbing its crisp white pages, but we will get used to it. I'm not sure that we should get used to it, because keystroke reference reduces the writer's mobility and courts the dangers of eye-strain, back-ache and repetitive strain injury.

I don't quarrel with Isaac Asimov's description of the ideal medium for information. It should, he thought, be optimally ordered, the medium should have a low energy requirement and should be portable. The acronym for that comes from Bound Optimally Ordered Knowledge — BOOK. For some reason we've given this the retronym 'hard copy'.

The Oxford Writer's Shelf, The Oxford Science Shelf, The Concise Oxford Dictionary, The Oxford English Dictionary (CD-Rom), The Oxford Thesaurus, Complete Writers' Toolkit, Lexica, Random House Encyclopedia, Quotemaster, Correct Quotes.

7. Stylistic Help

Finally, in my last band fall usage- and grammar-checking programs and style analysers, which must underly mechanical translation. Here, software linguists have serious difficulties in playing off real live understanding against computer simulation. Clearly computers can only look at the structure of language without grasping its meaning. Chomsky's famous colourless green ideas that sleep furiously would pass muster in any of the available checking packages. One must accept that computer products will be limited in their efficiency and use them as part of the process of critical reading. This is where the game factor comes in. It is more fun to sit clicking buttons, accepting suggested changes and ignoring inappropriate ones than to read yourself. Nothing *can* replace intelligent reading, but software may enliven the editing task.

The challenge for program developers is to discover how you isolate the circumstances under which a computer can offer reliable advice on style and usage. Presupposing a market largely made up of business users makes the task a little easier. It *is* possible where error-trapping is reasonably trivial, and this in itself is of value to a writer-editor whose attention may inevitably nod. Thus it is easy enough to net a phrase like 'of that ilk' and a reasonable guess to assume that the writer doesn't know what 'ilk' really means. Software limitations, as well as a preconceived notion of audience interest, require disposing of this in two lines, but Philip Howard, in *A Word in Time*, has three pages on it. Here, as in so many usages, it is becoming so universal in its 'incorrect' form that only purists will object.

Software linguists face a dilemma. Do you leave out all such usages so that the writer is never alerted to a possible area of concern? Or do you tuck it away with some bland expression like, 'Avoid this phrase in formal writing'? Software that upbraids a writer too often is irritating and saps the confidence of inexperienced writers: software that fails to deal with one of the matters that can expose a writer to public scorn is unreliable and saps the confidence of experienced writers.

I have yet to find a grammar-checker that really works, though some make brave attempts and some are available in dual language versions (e.g. checks your French grammar but gives explanations and tutorials in English). The writer at play does better with a concordance, though it is much harder work than flowing words through a grammar-checker because considerable thought goes into isolating what one wants to know. But it is an eye-opener to run a series of analyses and worth doing once in your writing life if only for the light it sheds on personal style: the variety of one's own vocabulary; its musical flow; whether similar sentence characteristics occur across all one's writing; what weaknesses the relentless counting of words reveals. Such personally tailored analyses are of value to translators or originators alike. A concordance program is one of the more interesting electronic games in the writer's playground but are hard work to operate.

Correct Grammar, Grammatik 5, StyleWriter, RightRighter, Macproof, TexAtrix, The Oxford Concordance Program, Docucomp, The Editor's Assistant, Corporate Voice, Readability Plus, Globalink.

8. Playing with Copies

None of these categories excludes the other and they fit into a large library of tools for writers. In an ideal joint simulation they would link to each other. All would have access to the depths of pages of Howard, Fowler, Burchfield and Quirk *et al* hidden within the computer's interior and providing discussions and cross-references at whatever level of detail required. Such a Fowler-in-a-box — to borrow Geoffrey Nunberg's phrase [Nunberg (6)] — would require hypertext on CD-Rom with all its attendant copyright problems. How easy it then would be to assemble compilations plundering from other writers' electronic commentaries.

For in that word 'copyright', which we have come to regard as the *raison d'etre* of all creative work, are a host of assumptions. At a basic level is the notion that a text (original writing or translation) is a 'copy' or saleable unit. And ephemeral as an electronic text may seem, it is still to be regarded in the same way as its 'hard' equivalent. Copyright exists as soon as you type the words even if you have the monitor switched off and cannot read them. (It doesn't exist in a second-language text that is an unaided translation by a computer program, but that is another issue.)

However, computer programs such as the ones under discussion encourage the user to play with electronic texts; to manipulate, change, check and analyse them. At the very least such manipulation is one way of animating another writer's characters on screen: should the user then be paying for film rights? This is a facetious point, but not as frivolous as it sounds, for the unfortunate truth is that there is no adequate provision for texts that exist in electronic

form. On-line databases and text storage archives do not have a collecting system in place whereby copyright holders are reimbursed for the copying on of their texts. The honour system still holds but whether it is useful in a world where the technology positively encourages people to patch, purloin and plagiarise is another matter.

This aside, the computer tools in existence make it very easy for writers to supply machine-readable texts ready for use in the publication process. But publishers have so far been slow to regard this as an advantage. Only a third of word-processing authors in a recent survey reported that their publishers were interested in using their disks [Dorner, (2)]. Moreover, the Minimum Terms Agreement negotiated with publishers by the Society of Authors and the Guild of Writers has no clauses recognising the author's right to offer to the publisher electronic copy, nor discusses division of responsibilities or circumstances of use of such electronic copy, nor the subsidiary rights attendant on electronic copies. These are large issues, worth more attention than there is time for here, and introduced to indicate how ill-equipped we are to make provision for this world of virtual writing.

And if this is a virtual world, why not add sound and moving pictures to pep it up? After all, virtual reality is an illusion, and if my parallel is to work then the audio-visual element is essential. But *is* my analogy a good one? I am myself divided. In some sense I feel that, limited though they are, the programs I have seen educate because they allow playtime with written language. The point-and-click magic of correcting sentences on the wing is both fun and face-saving. This is something we have been missing in library-book jeremiads. On the other hand today's commercial products do seem tame. Maybe this will change in a few years' time, when language corpora (such as COBUILD and Lancaster/Longman) have analysed hundreds of millions of words and phrases to give a databank of current English usage. Then — given adequate memory requirements — automatic parsers may begin to look sensible. Then translation may be made easier. Until that time we must remember that computers are playing in a world of virtual writing — very like the real thing, but without its substance.

References

1. Dorner, J. 1992, *Writing on Disk* London: John Taylor Book Ventures.
2. Dorner, J. 1991, *Authors and Information Technology: New Challenges in Publishing* BNBRF Report No. 52, London: The British Library.
3. Svartvik, J. (ed), 1992, *Directions in Corpus Linguistics* Berlin: de Gruyter.
4. Galbraith, D. 1991, 151-160 'Conditions for Discovery through Writing', *Computers and Writing IV*.
5. Masterman, M. in Reichhart, J. (ed) 1971, 175-83, 'Computerised Haiku' *Cybernetics, Art and Ideas*, London: Studio Vista.
6. Nunberg, G. 1990, 'Review of the Longman Guide to English Usage', *International Journal of Lexicography*, 3 (2), 111-132, Oxford: The University Press.

Parts of this paper are adapted from the author's 'Virtual English', 1992, *English Today*, 8 (4), Cambridge: The University Press.

On hard disk testing random doesn't matter, because every byte is written as is (also on ssd with dd). But if we test dedupped zfs pool with pure zero or random data, there is huge performance difference. It takes care of the caching, syncing, random data, random location on disk, small size updates, large updates, reads, writes, etc. Comparing a usbkey, a harddisk (rotary), a solid-state drive and a ram-based filesystem can be very informative for the newbie. I am looking to write a file filter for disk.sys that needs to attach to the creation process before the operating system gains access to the disk/partition. However, I only need to read/write one-two bytes in the MBR nothing more and I am not very clear on the ASM/C required to do this. Hopefully this example will clarify my intentions. I need to layer the file filter for disk.sys so that when the Disk class driver creates the device objects representing the raw disks (any). \