MARGARET MASTERMAN: HER LIFE AND WORK
(An Appreciation by an erstwhile colleague)

Frank Knowles

Aston University

I first met Margaret Masterman in 1973: the occasion - and how apt it is that I relate this here - was the first Aslib Informatics conference at Durham. I had, prior to that event, received a letter out of the blue from Kevin Jones, who asked me whether I might be able to give a talk about mathematical linguistics. I accepted the invitation but with some misgivings: my credentials were very meagre - not worth the paper they weren't written on, as the saying goes! I had an interest in quantitative aspects of language study which dated back to the time when I was given systematic instruction in cryptography during military service. I had also made it my business to follow, as a lecturer in the field of Russian language and linguistic studies, what is called in Russian 'inženernaja lingvistika', or 'engineering linguistics'. My academic work had brought me into personal contact with Soviet scholars working in the field of machine translation, such as Rozencvejg, Piotrovskij, Fitialov and Soboleva. Moreover, my occasional work as a Russian interpreter had brought me face to face with notable Soviet figures in the domain of information science, such as Čërnyj, Mixajlov and Moskovič. Lastly, I was at a stage in my own development which I might describe by saying that I had recently been awarded my 'programmer's wings'! So off I went to Durham.

I was stimulated and fascinated by Margaret Masterman's 'presence' at Informatics 1: I remember her paper and 'interventions' during the working sessions and I recall with much greater clarity how she came into the crowded conference bar on the first evening, easily captured rhetorical control of the proceedings and launched a brainstorming discussion which lasted until the early hours of the following morning! What struck me then, and on many subsequent occasions, was Margaret's encyclopaedic grasp of all the fundamental issues - resolved and unresolved - in what would now be called AI. I was deeply impressed by the ease with which she conversed not only with other philosophers, but with linguists, mathematicians, electrical engineers, information scientists, psychologists, and computer scientists. My subsequent move from Manchester to Birmingham made it easier for me to get across to Cambridge to see and talk to this remarkable woman, which I did as often as circumstances permitted.

Early in 1978 we both found ourselves - along with many other researchers from all over Europe - at a meeting in Luxembourg convened by the European Commission for the purposes of identifying both strategies and people to cope with the introduction of IT hardware and techniques into the world of translation within the various echelons of the European Commission. The

European Commission had recently purchased Peter Toma's SYSTRAN machine translation system, originally designed for 'try anything' translation from Russian into English. Some members of the European Commission's own staff were already deployed on metamorphosing SYSTRAN into an English-French system and, additionally, fine-tuning it. It is quite easy to imagine the discussion sessions at that huge meeting, with people all the time raising new points, discovering hidden variables, and attempting to achieve a consensus - or at least a running recapitulation! Margaret's contribution to that mammoth debate, be it in the form of questions or answers, was instrumental in advancing the whole issue *a* considerable distance. One very important outcome of that meeting, as far as the Cambridge Language Research Unit (CLRU) was concerned, was a grant from the European Commission for investigating SYSTRAN: was it already operating at or near the ceiling of its design limits or did it have development potential which had hardly been tapped at all? Put in a different and more exciting way - fully consonant with today's insights into expert systems - what needed to be done to unlock the expertise embedded in SYSTRAN over the years by many different people? How could the experiential skills of those people be turned into a descriptive, cognitive statement of worth which might allow the system to be sanitised, symmetrised, 'de-obfuscated' and optimised? Margaret's eventual Project Report could not, of course, provide complete answers to all these vital questions but it certainly addressed them all very fully. I remember Margaret and Bob Smith coming across to Birmingham to see me for a couple of days during August 1979. Bob and I took advantage of an under-utilised mainframe to study the SYSTRAN 'evidence': 100,000 IBM 370 Assembler commands! What we did was to write a program that accepted the SYSTRAN program's Assembler commands as textual data and then tried to spell out *en clair* what was happening in the MT program. Margaret was behind us all the time, breathing down our necks and telling us what she wanted to find and often did find: she simply wouldn't let us break off until she had either got her answers or new problems to reflect on. That was her style! In that project and in many others Margaret was trying to do something very important: to elicit and to elucidate - and somehow to compact - heuristics into analytical statements, to define methods of establishing an intellectual grasp of, and potentially control over, procedures, often very intricate, used by human beings in their various professional contexts. Today we call this 'expert systems research' and place a high premium on that activity.

Although her 'hands on' experience was, I believe, very limited, Margaret derived a pleasure that was intellectual, emotional and almost visceral from her dealings with computers. It warmed her heart to know that utterly diverse types of data can be coded in an electronically reliable and uniquely identifiable manner; also, she was thrilled by the facility of being able to submit such data to programs, to process it - whatever that term nay mean in the actuality of many different applications and contexts - and to output 'results' with a consistency and a speed which are far in excess of the capacity of the human brain aided by human motor skills. The use of the word 'data' above needs to be glossed: as we all know, there is a special and crucially important type of data without which computers would be virtually emasculated - data in the form of operational instructions which are intended to be applied to 'real' data, in the normal sense of that term. We usually call this feature the 'stored program' concept: if the

word 'program' designates, as one of its meanings, an ordered sequence of activities, viewable as an entity, then it is clear that in computer parlance such a definition holds good as well. Margaret knew as well as any of us that programmers much predict. If their programs are to be useful, then certain logical options which can be selected according to given criteria must be testable within the processing environment itself; that is, at a remove in time and space from the programmer's original conception. All this merely emphasised for her the need to analyse problems, to break them down into subproblems and discrete steps so as then to be able to synthesise solutions in such a way that no further human intervention is required. The human, preferably algorithmic thought-process and problem-solving strategy has been captured and can be perpetuated at will in a computational environment. Margaret never lost her sense of excitement about computers. Even towards the close of her life she derived deep satisfaction from the knowledge that traditional programming languages - often characterised as 'imperative' languages because the individual program statements give orders - were being complemented by languages of a more recent and novel design, such as PROLOG, which can be classified as 'declarative/interrogative'. The essence of programming in such languages is to establish a set of entities, to assign their attributes as appropriate, to declare relationships holding between the entities, and subsequently to interrogate the resulting environment in order to reveal truth-values in such microcosms and to elicit relationships not declared contiguously or not even declared at all, Margaret herself was no programmer, but I am sure I am right in claiming that she knew an awful lot more about what sort of intellectual activity programming is than many full-time programmers do.

Machine translation (MT) was a major professional preoccupation for Margaret. As a philosopher she saw, quite correctly, that the main effort had to be directed towards the development of methods for preserving meaning and safeguarding it against corruption during the transformation process: she was fascinated by this problem and we often talked about two particular axes which researchers and implementers have to align with each other. The first is perhaps best expressed by saying that in MT - and in natural language processing (NLP) generally - what cannot be computed has to be looked up, and vice versa. In some cases both options are available, each with its own pay-off or, conversely, overhead; in other instances only one option is realistic - but how can this be determined and reconciled with the abiding need to maintain identity of sense as between input and output? All of the various types of meaning have to be identifiable and transformable in MT: denotational, connotational, collocational, stylistic, rhetorical, and - not least - syntactic.

The other axis which claims attention - inside and outside MT - is one which can perhaps be likened to a magnetic bar with its characteristic poles: textual and lexical. One can choose between two stances: either textocentric or lexicocentric. The former postulates text as a concatenation of discrete elements which merge to produce *a* structure with a sense greater than the sum of the individual meanings of the component parts. Text in this view must be considered to be a gestalt. It must be perceived holistically and 'tampering' with it in any way can have disastrous consequences. Margaret was - not surprisingly, in view of her

education - ever aware of the etymological derivation of the word 'text', from Latin *textus,* designating 'something woven' (c.f. 'toga'). For her - and, I am sure, for all of us - the author of a text needs to wield his pen with as much care and skill as a weaver operates his loom! What is it that actually happens when lexical units are inserted into a textual environment? And what occurs when texts are decomposed into their constituent units? How can text be segmented so as to yield only elements which possess extra-textual meaning and which are, putting it another way, the stock-in-trade of lexicographers, espousers of the other, the lexicocentric view of things, the codifiers and custodians of discrete lexemes? For Margaret, lexicographers were kindred spirits: they were in the business of defining words, using, wherever possible, the classical method of defining *per genus et differentiam,* making use - but often only subliminally - of semantic primitives. Margaret could also criticise lexicographers for not using presentation methods which allow structural information to be made manifest. Most of them were hidebound, in Margaret's view, by their unquestioning allegiance to unintelligent listing methods such as alphabetic ordering, and by their overconcentration on the properties of members of lexical sets rather than on the properties of the sets themselves. Margaret could not understand why lexicographers so very often seemed to pass by the challenge of really getting into taxonomy and onomasiology. For her lexicographers had to be lexicologists as well, willing to look at new opportunities such as thesauri and especially computerised lexical databases.

Margaret had a strong interest in syntax: for her, syntax, that is surface syntax, represented a vital crux. What is the process by which atoms of meaning are configured into molecules and thence into whole structures and edifices of sense, into what Margaret was wont to call 'messages'? She was particularly preoccupied with 'differential syntax', the myriad of extant syntactic systems in the world's languages, all of them representing a different meld, however, of just three linguistic mechanisms: inflection, the use of function words, and element order. Why do the words of some languages parade themselves, as it were, in military uniforms, replete with badges to show what part of speech they are; whereas words in other languages are quite content to wear civilian clothes, to be gregarious, and ultimately to be judged by the company they keep? Of course, the basic building blocks of language do not operate in grand isolation of one another: the behaviour of one subsystem affects, and is affected by, its partners. Margaret saw this very clearly and had a very good grasp of many of the 'nuts and bolts' of syntax, for someone who was not a mainstream linguist involved in the detailed study of language systems all of the time. I remember arriving in Cambridge one evening, after a particularly fatiguing journey, to be greeted in the CLRU lobby by Margaret proclaiming: 'Do you realise that today's syntax is yesterday's morphology? This means that Chinese is the world's most highly evolved language because its morphology has been completely eroded away!' Margaret was particularly energetic in her quest for a satisfying personal view of the ways in which syntactic mechanisms interact with techniques for the thematic organisation of text. She could easily view subject and predicate both as a grammarian and as a logician. I know she dearly wanted to conduct extensive experimentation on a large corpus of text for the purposes of evaluating, both in qualitative and quantitative terms, theme-rheme behaviour. She

always hoped, to put it figuratively, that the syntacticians' stalagmite would meet and blend with the rhetoricians' stalactite! This urgent hope was the well-spring of her work on rhythm.

Margaret, furthermore, had a close concern with text and textuality. In a way she was a text linguist born and bred. She certainly explored cohesion, or the ways in which texts are internally glued together. Her primary focus was coherence, or what makes individual concatenated utterances into acceptable, large-scale bodies of information, not compromised by any failures or inconsistencies in presuppositions or entailments, enhanced by delicacy, and suitable deference to the needs of those whose task it is to comprehend them. Pre-eminent, for Margaret, among the various text modes was expository prose displaying clear argumentation, permitting easy inferencing, and not burdened with too much cyclicity or convolution. She had, however, an abiding love for text of a poetical nature, full of imagery and redolent with evocative meaning. Margaret saw all text and all discourse as examples of systems fashioned by human beings who are able to use mechanisms and devices to perfection without, apparently, feeling the compelling need she felt to probe into them and lay them bare. One particular type of text which fascinated Margaret was translated text: it follows that she was just as much interested in human translation as in MT. She viewed human translation as the acme of skilled linguistic activity, rating it higher - I am tempted to say - than original creative writing because the translator has the constraint of needing to fully reveal the brilliance of the original author's mind whilst totally concealing the brilliance of his own! However, Margaret was by no means oblivious to other highly skilled linguistic activities, such as paraphrasing, summarising, stylistic transposition, all of them backed up by vital subsystems, such as deixis, comparison, analogy, enumeration, exemplification, generalisation, ellipsis, etc. Margaret conceded readily that present-day MT is clumsy simulation relying on sleight of hand. It is not emulation. Even simulation practised by perfect prestidigitators would not have been sufficient for her: she wanted MT - and IT, for that matter - to actually emulate human behaviour. Nothing less would do!

My other major piece of collaboration with Margaret was on the CLRU 'breathgroup' project, so deeply and positively influenced throughout, and so ably and energetically brought to completion, by Bill Williams after the onset of Margaret's final illness. Margaret's starting-point on this project was one which was typical of her: the postulation of a linguistic universal: to wit, that when human beings speak they draw - in normal circumstances - new breath only at positions which represent proper and distinct transitions in the flow of ideas. Any study of breathgroups should hence show logical 'stepping stones' in the elaboration of people's utterances. There was a time in history when texts were written down chiefly because it was known that they would need to be declaimed later. (How interested Margaret was, incidentally, in the 'orality and literacy' debate and its implications for IT!) Punctuation is hence a (primitive) system of suggesting to the orator how he should juncture his delivery. In fact, 'period' (usually known as 'full-stop' in British English) is a very clear piece of terminology: less clear but equally informative are terms such as 'comma', meaning 'cut', and 'colon', meaning 'breathgroup'! Lest any misunderstandings arise, it should be made clear that the CLRU breathgroup

project was not concerned with phonology or speech analysis but with written text. The connection is subtle but revealing: Margaret vehemently insisted that any writer should always read aloud what he or she had written, at least in front of a bathroom mirror and preferably on a dais in front of five hundred people, without a microphone! This was a sure method, she claimed, of helping writers to determine whether they had anything useful to say, whether they had got their ideas in the right order, and whether their message was likely to carry conviction. Writers are therefore always orators. Good writers do not need to be told this: being an orator is second nature to them. It follows that if the output of good writers is studied in its written form something valuable might be gleaned about the way in which ideas, or cognitive units, interrelate. The actual methods of conducting such research go beyond the scope of this commemorative paper - a brief indication must suffice: an algorithm, capable of extensive pattern-matching, has to be constructed which computes, on the basis of data on punctuation, stress, parts of speech, etc., where cuts in the enunciation need to be made. (Texts may be rhetorically punctuated in this fashion, requiring speakers - such as orators, clergymen or news-readers - to draw one breath per line of text.) The resulting segments then need to be studied, 'homogenised', refined, classified and somehow made amenable to lexicographical treatment. We can only hope that Margaret was somehow able to appreciate, in the terminal phase of her distressing illness, that the work done during the three-year period of the breathgroup project amply justified her and others' confidence in such research.

I conclude by stating what a wonderfully stimulating person Margaret was to know and to work with. She had her idiosyncrasies - but so do we all! For her, the message was what was important - not necessarily the medium or indeed the messenger. The tribute I offer her is quite simply to say that, in my view, she was way ahead of her time. On a more personal note I quite freely and openly acknowledge that, as a philosopher, she taught me, a linguist, to think about things about which I should already have been thinking but about which I had been blithely ignorant - I shall be ever grateful to her for the service she thereby rendered me.

Requiescat in pace!