Editor's introduction

1. A personal memoir: Margaret Masterman (1910–1986)

Margaret Masterman was ahead of her time by some twenty years: many of her beliefs and proposals for language processing by computer have now become part of the common stock of ideas in the artificial intelligence (AI) and machine translation (MT) fields. She was never able to lay adequate claim to them because they were unacceptable when she published them, and so when they were written up later by her students or independently 'discovered' by others, there was no trace back to her, especially in these fields where little or nothing over ten years old is ever reread. Part of the problem, though, lay in herself: she wrote too well, which is always suspicious in technological areas. Again, she was a pupil of Wittgenstein, and a proper, if eccentric, part of the whole Cambridge analytical movement in philosophy, which meant that it was always easier and more elegant to dissect someone else's ideas than to set out one's own in a clear way. She therefore found her own critical articles being reprinted (e.g. chapter 11, below) but not the work she really cared about: her theories of language structure and processing.

The core of her beliefs about language processing was that it must reflect the coherence of language, its redundancy as a signal. This idea was a partial inheritance from the old 'information theoretic' view of language: for her, it meant that processes analysing language must take into account its repetitive and redundant structures, and that a writer goes on saying the same thing again and again in different ways; only if the writer does that can the ambiguities be removed from the signal. This sometimes led her to overemphasise the real and explicit redundancy she would find in rhythmical and repetitive verse and claim, implausibly, that normal English was just like that if only we could see it right.

This led in later years to the key role she assigned to rhythm, stress, breath groupings and the boundaries they impose on text and the processes of understanding. To put it crudely, her claim was that languages are the way they are, at least in part, because they are produced by creatures that

2 Editor's introduction

breathe at fairly regular intervals. It will be obvious why such claims could not even be entertained while Chomsky's views were pre-eminent in language studies. But she could never give systematic surface criteria by which the breathgroups and stress patterns were to be identified by surface cues, or could be reduced to other criteria such as syntax or morphology, nor would she become involved in the actual physics of voice patterns.

Her views on the importance of semantics in language processing (which she continued to defend in the high years of Chomskyan syntax between 1951 and 1966) were much influenced by Richens' views on classification and description by means of a language of semantic primitives with its own syntax. These, along with associated claims about semantic pattern matching onto surface text, were developed in actual programs, from which it might be assumed that she was a straightforward believer in the existence of semantic primitives in some Katzian or Schankian sense. Nothing could be further from the truth: for she was far too much a Wittgensteinian sceptic about the ability of any limited sublanguage or logic to take on the role of the whole language. She always argued that semantic primitives would only make sense if there were empirical criteria for their discovery and a theory that allowed for the fact that they, too, would develop exactly the polysemy of any higher or natural language; and she always emphasised the functional role of primitives in, for example, resolving sense ambiguity and as an interlingua for MT.

She hoped that the escape from the problem of the origin of semantic primitives would lie in either empirical classification procedures operating on actual texts (in the way some now speak of deriving primitives by massive connectionist learning), or by having an adequate formal theory of the structure of thesauri, which she believed to make explicit certain underlying structures of the semantic relations in a natural language: a theory such that 'primitives' would emerge naturally as the organising classification of thesauri. For some years, she and colleagues explored lattice theory as the underlying formal structure of such thesauri.

Two other concerns that went through her intellectual life owe much to the period when Michael Halliday, as the University Lecturer in Chinese at Cambridge, was a colleague at CLRU. She got from him the idea that syntactic theory was fundamentally semantic or pragmatic, in either its categories and their fundamental definition, or in terms of the role of syntax as an organising principle for semantic information. She was the first AI researcher to be influenced by Halliday, long before Winograd and Mann. Again, she became preoccupied for a considerable period with the nature and function of Chinese ideograms, because she felt they clarified in an empirical way problems that Wittgenstein had wrestled with in his so-called picture-theory-of-truth. This led her to exaggerate

Editor's introduction

the generality of ideogrammatic principles and to seem to hold that English was really rather like Chinese if only seen correctly, with its meaning atoms, being highly ambiguous and virtually uninflected. It was a view that found little or no sympathy in the dominant linguistic or computational currents of the time.

Her main creation, one that endured for twenty years, was the Cambridge Language Research Unit, which grew out of an informal discussion group with a very heterogeneous membership interested in language from philosophical and computational points of view. Subsequently, the attempt to build language-processing programs that had a sound philosophical basis was a distinctive feature of the unit's work. This approach to language processing, and the specific form it took in the use of a thesaurus as the main vehicle for semantic operations, will probably come to be seen as the unit's major contributions to the field as a whole, and it was Margaret who was primarily responsible for them. Her vision of language processing and its possibilities was remarkable at a time when computers were very rudimentary: indeed, much of the CLRU's work had to be done on the predecessors of computers, namely Hollerith punched-card machines. Equally, Margaret's determination in establishing and maintaining the unit, with the enormous effort in fundraising that this involved, was very striking: the fact that it could continue for decades, and through periods when public support for such work was hard to come by, is a tribute to Margaret's persistence and charm. It is difficult for us now, in these days of artificial intelligence in the ordinary market-place, and very powerful personal computers, to realise how hard it was to get the financial resources needed for language-processing research, and the technical resources to do actual experiments.

Perhaps the best comment on Margaret's initiative in embarking on language-processing research, and specifically on machine-translation work, comes from a somewhat unexpected source. Machine translation, after an initial period of high hopes, and some large claims, was cast into outer darkness in 1966 by funding agencies who saw little return for their money. Reviewing twenty-five years of artificial-intelligence research in his presidential address to the American Association for Artificial Intelligence in 1985, Woody Bledsoe, one of the longstanding leaders of the field, though in areas quite outside language, said of those who attempted machine translation in the fifties and sixties: 'They may have failed, but they were right to try; we have learned so much from their attempts to do something so difficult'.

What MMB and CLRU were trying to do was far ahead of their time. Efforts were made to tackle fundamental problems with the computers of the day that had the capacity of a modern digital wristwatch. Despite every

4 Editor's introduction

kind of problem, the unit produced numerous publications on language and related subjects, including information retrieval and automatic classification. For over ten years the unit's presence was strongly felt in the field, always with an emphasis on basic semantic problems of language understanding. Margaret had no time for those who felt that all that needed doing was syntactic parsing, or that complete parsing was necessary before you did anything else. Now that the semantics of language are regarded as a basic part of its understanding by machine, the ideas of CLRU seem curiously modern.

Margaret's main contribution to the life of CLRU was in the continual intellectual stimulus she gave to its research, and through this to the larger natural language processing community: she had wide-ranging concerns, and lateral ideas, which led her, for example, to propose the thesaurus as a means of carrying out many distinct language-processing tasks, like indexing and translation. Margaret's emphasis on algorithms, and on testing them, was vital for the development of CLRU's work on language processing; but her ideas were notable, especially for those who worked with her, not just for their intellectual qualities, but for their sheer joyousness.

Her colleagues and students will remember her for her inspiration, rather than her written papers: she made questions of philosophy and language processing seem closely related and, above all, desperately important. On their joint solutions hung the solutions of a range of old and serious questions about life and the universe. In this, as so much else, she was a Wittgensteinian but, unlike him, she was optimistic and believed that, with the aid of the digital computer, they could be solved.

She could not only inspire and create, but terrify and destroy: she had something of the dual aspects of Shiva, an analogy she would have appreciated. Even in her seventies, and still funded by European Commission grants, her hair still black because a gypsy had told her forty years before that it would not go grey if she never washed it, she would rise, slowly and massively at the end of someone's lecture, bulky in her big, belted fisherman's pullover, to attack the speaker, who would be quaking if he had any idea what might be coming. The attack often began softly and slowly, dovelike and gentle, gathering speed and roughness as it went. As some readers may remember, there was no knowing where it would lead.

2. Themes in the work of Margaret Masterman

In this introductory chapter I shall seek to reintroduce and then focus the work of Margaret Masterman by enumerating and commenting briefly on a number of themes in her work. Some of these have been successful, in the sense of appearing, usually rediscovered, in some established place in the

Editor's introduction

field of natural language processing, while others, it must be said, appear to have failed, even though they remain highly interesting. This last is a dangerous claim of course, one that can be reversed at any time. There is in my view a third category, of general programmes rather than particular representational methods, about which one can only say that they remain unproven. In spite of their breadth, scope and originality it must also be conceded that Margaret Masterman did not have theories to cover all aspects of what would be considered the core issues of computational linguistics today: for example, she had little or nothing to say on what would now be called text theory or pragmatics. Nor did she have any particular reason for ignoring them, other than that she thought the problems that she chose to work on were in some sense the most fundamental.

The order of the themes corresponds broadly to that of the sections of this book: it moves from abstract concepts towards more specific applications of those concepts, from particular forms to language itself, on which those forms imposed the coherence and redundancy that she believed to be at the core of the very idea of language. I shall continue here the affectionate tradition of referring to her as MMB, the initials of her married name Margaret Masterman Braithwaite.

2.1. Ideograms

This was an early interest of MMB's (Masterman, 1954 and Chapter 1) that persisted throughout her intellectual life: the notion that ideograms were a fundamental form of language and were of non-arbitrary interpretation. The root of this idea lay in Wittgenstein's interest (1922) in how pictures could communicate: in how the drawing of an arrow could convey movement or pointing and, before that, in his so-called Picture Theory of Truth, where objects could be arranged to express facts. More particularly, she must almost certainly have been influenced by his Notebooks 1914–1916, where he writes, 'Let us think of hieroglyphic writing in which each word is a representation of what it stands for'.

The connection of all this to ideograms had been noted by I.A. Richards, who was much preoccupied by Chinese, and who developed *English Through Pictures* (Richards and Gibson, 1956), a highly successful language-teaching tool. MMB came to Chinese through Michael Halliday, then a Cambridge University lecturer in Chinese, and began to use stick-pictures as representations of situations which could also provide a plausible referential underpinning for language: something universal, and outside the world of the language signs themselves, yet which did not fall back on the naive referentialism of those who said that the meanings of words were things or inexpressible concepts.

5

6 Editor's introduction

Frege (new translation, 1960) had tackled this issue long before and created a notation in which propositions had a sense, but could only refer to the true or the false (at which point all differences between them, except truth value, were lost). This reference to situations, which MMB helped keep alive, has found formal expression again in Barwise and Perry's *Situation Semantics* (1983). They, too, wanted a central notion of a situation as what an utterance points to, and they too resort to cartoon-like pictures but, unlike MMB, nowhere acknowledge the role of Wittgenstein's Picture Theory of Truth.

It is as hard to capture the future in this field as of any other, and the movement of a (partially) ideogrammatical language like Japanese to centre stage in language processing may yet show the importance of ideograms for grasping the nature of language. But whatever is the case there, MMB's interest remained not only in the differences in the ways occidental and the main oriental language represent the world, but also in the ways those differences reflect or condition basic thought: she liked to quote a phrase of Whitehead's that our logic would have been better based on the Chinese than the Greeks.

2.2. Lattices and Fans

Although not a formalist herself, and considered an anti-formalist by many, MMB nevertheless believed passionately in the applicability of mathematical techniques to natural language; without them, she believed, there would be nothing worthy of the name of theory or science. What she was opposed to was the assumption that formal logic, in particular, could be applied directly to natural language, and she would not concede much distinction between that and the methods of Chomsky (1965), a position that has some historical justification.

The two structures from which she hoped for most were lattices and 'fans', a notion she derived from some work of Brouwer (1952). MMB believed lattices (Masterman, 1959 and Chapter 3) to be the underlying structure of thesauri, and fans (Masterman, 1957b and Chapter 2), she believed, mapped the spreading out of the new senses of words, indefinitely into the future. She spent some time trying to amalgamate both representations into a single structure. These efforts have not met with much success nor have they been taken up by others, although Zellig Harris did at one time toy with lattices as language structures, and Mellish (1988) has sought to link lattice structures again to Halliday's categories of grammar and semantics.

Another problem is that fans are too simple to capture much: they have no recursive structure. And lattices are so restrictive: once it is conceded

Editor's introduction

that neither words nor things fall neatly under a taxonomic tree structure, it by no means follows that they fall under a graph as restricted as a lattice either. More promising routes have been found through more general applications of the theory of graphs, where the constraints on possible structures can be determined empirically rather than a priori.

2.3. Thesauri and the use of large-scale language resources

MMB believed thirty years ago that constructed entities like dictionaries and thesauri (especially the latter) constituted real resources for computational language processing (Masterman, 1956, 1958 and Chapters 4 and 6, respectively). That was at a time when any computational operations on such entities were often dismissed, by those working in other areas of computational linguistics, as low-grade concordance work. Betty May compacted the whole of Roget's Thesaurus for MMB, from a thousand 'heads' to eight hundred, and had it put onto punched cards. That formed the basis for a range of experiments on Hollerith sorting machines, which contributed to Karen Spärck Jones' seminal thesis work Synonymy and Semantic Classification (1964, 1986). MMB believed that thesauri like Roget were not just fallible human constructs but real resources with some mathematical structure that was also a guide to the structures with which humans process language. She would often refer to 'Roget's unconscious' by which she meant that the patterns of cross-references, from word to word across the thesaurus, had generalisations and patterns underlying them.

In recent years there has been a revival of interest in computational lexicography that has fulfilled some of MMB's hopes and dreams. It has been driven to some extent by the availability from publishers of machine-readable English Dictionaries, like LDOCE and COBUILD, with their definitions written in a semi-formal way, one that makes it much easier for a computational parser to extract information from them. But the initial work in the current wave was done by Amsler (1980) at Texas using Webster's, an old-fashioned dinosaur of a dictionary. He developed a notion of 'tangled hierarchies', which captures the notion MMB promoted to get away from straightforward tree-like hierarchies.

Current centres for such work include Cambridge, Bellcore, IBM-New York, Waterloo, Sheffield and New Mexico, where it has been carried out by a number of techniques, including searching for taxonomic structures, by parsing the English definitions in the dictionary entries, and by collocational techniques applied to the word occurrences in the entries themselves. This last normally involves the construction in a computer of very large matrices, as foreseen in the earlier work of Spärck Jones. Those matrices

7

8 Editor's introduction

can now be computed effectively with modern machines in a way that was virtually impossible twenty-five years ago.

Although dictionaries and thesauri are in some sense inverses of each other, they also differ importantly in that dictionaries are written in words that are themselves sense-ambiguous, except, that is, for those entries in a dictionary that are written as lists of semi-synonyms (as when, for example, 'gorse' is defined as 'furze' and vice-versa). One of the major barriers to the use of machine-readable dictionaries has been the need to resolve those lexical ambiguities as the dictionary itself is parsed, which is to say, transformed by computer into some more formal, tractable structure. MMB was more concerned with thesauri than dictionaries as practical and intellectual tools, and they do not suffer from the problem in the same way. Words in a thesaurus are also ambiguous items, but their method of placement determines their sense in a clearer way than in a dictionary: the item 'crane', for example, appears in a thesaurus in a list of machines, and therefore means a machine at that point and not a bird. The name 'machine' at the head of the section can thus straightforwardly determine the sense of items in it. Yarowsky (1992) returned to Roget as a basis for his fundamental work on large-scale word sense discrimination.

However, the last ten years have seen the Princeton WordNet (Miller et al. 1990) take over from dictionaries like LDOCE as the most-used linguistic-semantic resource. WordNet is a classic thesaurus, made up from scratch but with a powerful indexing mechanism and a skeletal set of categories and relations replacing the Roget 1,000 heads.

2.4. The use of interlinguas

MMB was much associated with the use of interlinguas (or universal languages for coding meaning) for MT and meaning representation (Masterman, 1967 and Chapter 7), and her reply to Bar-Hillel's criticism (1953) of their use has been much quoted. The notion of a uniform and universal meaning representation for translating between languages has continued to be a strategy within the field: it had a significant role in AI systems like conceptual dependency (Schank 1975) and preference semantics (Wilks 1975a), and is now to be found in recent attempts to use Esperanto as an interlingua for MT.

MMB's own view was heavily influenced by the interlingua NUDE (for naked ideas or the bare essentials of language) first created by R. H. Richens at Cambridge for plant biology: in a revised form it became the interlingua with which CLRU experimented. NUDE had recursively constructed bracketed formulas made up from an inventory of semantic primitives, and the formulas expressed the meaning of word senses on

Editor's introduction

English. Karen Spärck Jones worked on making NUDE formulas less informal, and defining the syntactic form of those entries was one of my own earliest efforts, so that a revised form of NUDE became my representational system for some years. In that system some of Richens' more 'prepositional' primitives had their function merged with what were later to become case labels, in the sense of Fillmore's Case Grammar (1968), for example, Richens' TO primitive functioned very much like Fillmore's Destination Case.

However, MMB's attitude to these primitives was very unlike that of other advocates of conceptual primitives or languages of thought: at no point did she suggest, in that way that became fashionable later in Cognitive Science, that the primitive names constituted some sort of language in the mind or brain (Fodor's view, 1975) or that, although they appeared to be English, the primitives like MOVE and DO were 'really' the names of underlying entities that were not in any particular language at all. This kind of naive imperialism of English has been the bane of linguistics for many years, and shows, by contrast, the far greater sophistication of the structuralism that preceded it.

MMB was far too much the Wittgensteinian for any such defence of primitive entities, in this as in other matters: for her, one could make up tiny toy languages to one's heart's content (and NUDE was exactly a toy language of 100 words) but one must never take one's language game totally seriously (linguists forgot this rule). So, for her, NUDE remained a language, with all the features of a natural one like English or French, such as the extensibility of sense already discussed.

That tactic avoided all the problems of how you justify the items and structure of a special interlingual language that are claimed to be universal, or brain-embedded, of course, but produced its own problems such as that of what one has achieved by reducing one natural language to another, albeit a smaller and more regular one. This, of course, is exactly the question to be asked of the group proposing Esperanto as an interlingua for MT.

She would put such questions forcefully to those in CLRU who showed any sign of actually believing in NUDE as having any special properties over and above those of ordinary languages, a possibility she had herself certainly entertained: this was the technique of permanent cultural revolution within an organisation, known to Zen Buddhists, and later perfected by Mao Tse-tung.

MMB believed that such interlinguas were in need of some form of empirical justification and could not be treated as unprovable and arbitrary assumptions for a system, in the way Katz (1972) had tried to do by arguing by analogy from the role of assumed 'axiomatic' entities in physics

10 Editor's introduction

like photons or neutrons. One weak form of empirical support that was available was the fact that statistics derived from dictionaries showed that the commonest defining words in English dictionaries (exempting 'a' and 'the' and other such words) corresponded very closely indeed for the first 100 or so items to the primitives of NUDE. But MMB wanted something more structural than this and spent considerable time trying to associate the NUDE elements with the classifying principles of the thesaurus itself, which would then link back to the distributional facts about texts that the thesaurus itself represented. In this, as in other ways, MMB had more intuitive sympathy with the earlier distributional or structural linguistics, like Zelig Harris, than with the more apparently mathematical and symbolic linguistics of Chomsky and his followers.

2.5. The centrality of machine translation as a task

There is no doubt that MT has become in recent years a solvable task, at least for some well-specified needs, sometimes by the use of new representational theories, but more usually by means of better software engineering techniques applied to the old methods. Merely doing that has yielded better results than could have been dreamed of two decades ago.

MMB must be credited with helping to keep belief in MT alive during long years of public scepticism, and above all with the belief that MT was an intellectually challenging and interesting task (Masterman, 1967, 1961; Chapters 6 and 8, respectively). I think that is now widely granted, although it was not conceded within artificial intelligence, for example, until relatively recently. There it was still believed that, although language understanding required inference, knowledge of the world and processing of almost arbitrary complexity, MT did not: for it was a task that required only superficial processing of language. I think that almost everyone now concedes that that view is false.

What MMB sought was a compromise system of meaning representation for MT: one that was fundamental to the process of translation, but did not constitute a detailed representation of all the relevant knowledge of the world. She believed there was a level of representation, linguistic if you will, probably vague as well, but that was sufficient for MT and, in that sense, she totally denied the assumption behind Bar-Hillel's (1953) critique of MT, and which was taken up by some artificial intelligence researchers afterwards (though not, of course, the same ones as referred to in the last paragraph), that MT and language understanding in general did require the explicit representation of all world knowledge. This position of hers cannot be separated from her quasi-idealist belief that world knowledge cannot be represented independently of some language, and hence any true