Léon J., 2004, *" The inkstand was in the pen* and other stories. The controversy between Bar-Hillel and the Cambridge Language Research Unit about language formalization and machine translation " *Bulletin of the Henry Sweet Society* n°42 :4-10.

Jacqueline Léon, CNRS, Université Paris 7, France
jacqueline.leon@linguist.jussieu.fr

The inkstand was in the pen and other stories. The controversy between Bar-Hillel and the Cambridge Language Research Unit about language formalization and machine translation.

In the late 1950s, the first experiments in Machine Translation (MT below) gave rise to a controversy which opposed two conceptions of language and language formalization. The first one, claimed by Yehoshua Bar-Hillel, was a mathematico-logical view of language. The second one, advocated by Margaret Masterman and the Cambridge Language Research Unit, was grounded on meaning in context. This argument was focussed on the feasibility of MT, and especially the translation of sentences like *the pen was in the box/ the box was in the pen*, or *the pen was in the inkstand / the inkstand was in the pen*. Two reports help tracing this controversy : Bar-Hillel's " The present status of Automatic Translation of Language " (1959, published in 1960) and the CLRU's *Essays on and in Machine Translation* (unpublished, 1959), dedicated to Bar-Hillel in response to his report.

Yehoshua Bar-Hillel (1915-1975) and Margaret Masterman (1910-1986) were both MT pioneers. Bar-Hillel, an Israeli philosopher and logician, obtained a scholarship in 1950 to do his PhD with Rudolf Carnap (1891-1970) at the University of Chicago. In 1951, he was engaged as a full-time MT researcher (the first in the world) by MIT, where he organized the first colloquium on MT in 1952. He is the author of two reports on Machine Translation in 1952 and 1959. The second " The present status of Automatic Translation of Language ", published in 1960, rested on the evaluation of the first twenty MT centers all over the world. It contained serious criticisms of MT which led to its breakdown with the publication of the report of the Automatic Language Processing Advisory Committee (ALPAC) in 1966.

As early as 1952, Bar-Hillel was very critical towards the feasibility of FAHQT (Full Automatic High Quality Translation) and advocated computer-aided translation instead. In addition, he was interested in the formalization of natural languages and its

validation. He was especially convinced of the necessity of building an operational syntax for Machine Translation. Thus, in his 1953 paper entitled "A Quasi-Arithmetical Notation for Syntactic Description", he presented a method of syntactic description combining Ajdukiewicz' and Harris' works, in which a simple rule of quasi-arithmetical character is enough to compute the syntactic character of any linguistic string in its context, and provided the constituent structure of any given sentence mechanically. It should be said that Bar-Hillel's approach was very new among linguists: his operational syntax was the first categorial grammar as well as the first automatized syntactic parser. After a debate with Noam Chomsky (b.1928) about the place of logic in the study of language and especially the notion of transformation in the journal *Language* in 1954 and 1955, Bar-Hillel finally adopted Chomsky's transformational grammar which he thought should supplement the immediate constituent model to deal with *any* sentences, and not only simple (kernel) sentences (see Appendix II of his 1960 report).

Masterman founded the Cambridge Language Research Unit (CLRU) in 1955 to start experiments in Machine Translation in Britain. The group gathered many different and remarkable figures : Masterman was a philosopher, one of Ludwig Wittgenstein's (1889-1951) pupils, and married to Richard Braithwaite, professor of Moral Philosophy. R.H. Richens (-1984) was a biologist specialising in plant genetics. There also was linguists such as Martin Kay (b. 1935) and M[ichael] A[lexander] K[irkwood] Halliday (b.1925), and computer scientists, among them Yorick Wilks (b.1939), who became one of the first researchers on Natural Language Understanding. The originality of the Cambridge Unit is that it is the only MT group, besides the Russians, to develop a method of Machine Translation using intermediary language, that is a method founded on semantic representations common to every language[1]. The attention given to meaning transfer in Machine Translation was very uncommon among MT pioneers who, most of them, in the context of American structuralism, thought morphology and syntax, and in no way semantics, were dominant in the process of MT. This was also Bar-Hillel's view : syntactic analysis must have priority over semantics in the process of machine translation.

The controversy which opposed Bar-Hillel to the CLRU concerned the feasibility of MT, and especially the reduction of semantic ambiguities which, it must be added, is still a tricky issue for Natural Language Processing. In Appendix III of his 1960 report, Bar-Hillel's main argument against FAHQT (Full Automatic High Quality Translation) was that human translators appealed to extra-linguistic knowledge which machines

---

[1] On a survey of the CLRU's works, see Léon (2000).

cannot mobilize. To demonstrate this point, he invented the sentence *the box was in the pen*, taken from the linguistic context *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.*

> Assume, for simplicity's sake, that *pen* in English has only the following two meanings: (1) a certain writing ustensil, (2) an enclosure where small children can play. I now claim that no existing or imaginable program will enable an electronic computer to determine that the word *pen* in the given sentence within the given context has the second of the above meanings, whereas every reader with a sufficient knowledge of English will do this "automatically". (Bar-Hillel 1960 :159).

To solve this ambiguity, a context whatever it is, a paragraph or even a whole book, is of no use. Common knowledge is needed concerning the relative sizes of writing pens, toy boxes and playpens. This kind of extra-linguistic knowledge, as well as inferences, is not at the disposal of computers and no programs for the elimination of polysemy can deal with this kind of ambiguity[2]. Discussing the methods of reducing semantic ambiguity, Bar-Hillel regarded the use of the immediate linguistic environment as limited, and vehemently criticizes the CLRU method associating this method with a thesaurus : " Notice, e.g., that the very same − fictitious ! − thesaurus approach for English-to-French translation that would correctly render *pen* by " plume " in the sentence *The pen was in the inkstand* would incorrectly render *pen* by " plume " in the sentence *The inkstand was in the pen.* " (Bar-Hillel 1960 :162).

The CLRU members had a version of Bar-Hillel's report as early as February 1959 and they replied to Bar-Hillel's criticisms in a report *Essays on and in Machine Translation by the Cambridge Language Research Unit, dedicated to Yehoshua Bar-Hillel*[3]. The "pen was in the box " issue was dealt with in Essay ML91. It should be remembered that the very survival of the MT research group was at stake in this response.

Masterman claimed that semantics was fundamental in the process of machine translation. Thus the CLRU developed several intermediary language schemes from semantic representations common to every language[4]. One of them was based on thesaurus organization. Following Wittgenstein's view that a language is primarily a totality of contexts, not of sentences or words, and that the logical units for studying

---

[2] Only recently have Artificial Intelligence and Natural Language Understanding taken hold of this issue. See Winograd (1972) and his SHRDLU system for dealing with cubes of different sizes.

[3] An earlier discussion between Bar-Hillel and the CLRU members took place in November 1958 in Washington D.C., within the context of Bar-Hillel's evaluation.

[4] Some of them, such as the universal language Nude, conceived by Richens, was directly inspired by 17th century universal language schemes : Timothy Bright'Characterie' (1588), Wilkins' Essay (1668) and probably also by Dalgarno's Ars Signorum (1661), see Léon (2002).

language should not be words nor propositions but word contexts, namely word uses, Masterman wrote:

> The *Use* of a word is its whole field of meaning, its total "spread". Its usages, or main meanings in its most frequently found contexts, together make up its Use. (Masterman 1954 : 209)

Word meaning is thus meaning in context. And context refers to word use and also to the linguistic environment, the text, where the word occurs.

Because of its structure, based on the classification of words according to a set of contexts, Masterman chose thesaurus organization as a means of creating a new intermediary language, "a thesauric interlingua". At first, while pointing out its drawbacks, such as incoherence and non-systematicity, the CLRU chose Roget's Thesaurus.

In Essay ML91, Appendix II, Wordley presented a way of translating the two sentences *The pen was in the inkstand /The inkstand was in the pen* by establishing the meaning of the word *pen* with the thesaurus method. The heads of the words, *pen, in* and *inkstand* in Roget's Thesaurus are the following:

Pen: region, inclosure, limit, writing, book, hindrance, restraint, prison

In: existence, intrinsically, completeness, component, inclusion, contents, receptacle, nearness, contiguity, interiority, centrality, investment interjacence, inclosure, concavity, direction, approach, convergence, arrival, ingress, reception, insertion, qualification, meaning, method, importance, conduct, restraint, prison

Inkstand: permanence, receptacle, writing

The intersections between the three lists yield the following results :

pen inter *in*	inclosure, restraint, prison
pen inter *inkstand*	writing
in inter *inkstand*	receptacle

From these results, it may be observed that both the "plume" and the "enclosure" meanings are possible for both sentences. Thanks to a bracketing procedure, the analysis into immediate constituents and subject/predicate is carried out :

((the pen) / (is (in (the inkstand))))

((the inkstand) / (is (in (the pen))))

In the first sentence *pen* cannot be intersected with any other constituents in the sentence. No solution can be obtained and all the meanings of pen are kept. A larger context would be needed to obtain the "plume" meaning of *pen*. In the second sentence, *pen* can be intersected with *in* in the same constituent bracketing. This intersection restricts the meaning of *pen* to *inclosure, restraint*, and *prison*. The correct meaning of *pen* is then obtained.

Actually the issue went far beyond the reduction of semantic ambiguities in the MT process. More generally, in Essay ML91, named " Fictitious sentences in language " responding to Bar-Hillel's mocking expression " fictitious thesaurus " (cf. quotation 1960 :162, above), Masterman pointed out a fundamental difference of views between Bar-Hillel's philosophy of language and the CLRU's. She reproached Bar-Hillel for having based his argumentation against FAHQMT on a transposition of common sentences in English (*the pen was in the box, the pen was in the inkstand*) into " trick " ones (*the box was in the pen, the insktand was in the pen*), in addition to having ignored the possibility of there being a thesaurus entry for the word *in*, and the existence of a clause-bracketing program in the thesaurus procedure.

By " trick sentences ", Masterman referred on the one hand to sentences which were inflicted upon any MT researchers by the press at that time. Thus 'the whiskey was good but the meat had gone bad' was said to be the machine translation into Russian of 'the spirit is willing but the flesh is weak' ; likewise 'the invisible man is always insane' was said to be the machine translation into Japanese of 'out of sight, out of mind'. On the other hand, she named them " trick sentences " because they were provided as counter-examples by Bar-Hillel without any context. She completely disagreed with him on the issue of context, which she regarded as crucial in the determination of word meaning. She pointed out that, while he provided a context for *the box was in the pen*, that is *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy*, he gave no context for *the inkstand was in the pen*[5]. Without context, these trick sentences fail under a human translatability test : it can be shown that an ordinary American or English speaker fails to understand them.

Masterman provided a set of conditions which should be required of a counter-example to the possibility of FAHQMT (ML91 :5). It must come out of an actually occurrent context ; it must be shown accompanied by at least a page of that naturally occurrent context ; it must pass a human translatability test. Since they embody new contexts, these sentences are not 'fictitious' in ordinary language.

For Masterman, the issue of the possibility of MT depended in the end on which was right of two conflicting philosophies of language : Carnap's *Logical Syntax of Language*, supported by Bar-Hillel, or Wittgenstein's *Philosophical Investigations*. Both views of language, however, derived indirectly from Wittgenstein ; the first from his *Tractatus Logico-Philosophicus* (1922), the second from his *Philosophical Investigations* (1953).

---

[5] Here, however, the use of the preterite in this sentence could already be considered a kind of context because it inserts the sentence in a story.

In her view, Bar-Hillel substituted " normal order " sentences (the trick sentences) for normal English ones. Doing this, he adopted the mathematical conception of " normal order " used in Combinatory Logic. These are well-formed formulae - that is, meaningful sentences which are permutations, or bracketings, or re-operations, or repetitions of sentences in normal English. All such operations presuppose a view of language according to which the ordering of the total set of all sentences within such a language (within combinatory English, but not within normal English) can be defined by using calculus.

In contrast, Masterman claimed a philosophy of ordinary language where "sentences in normal English" are "sentences in English as English is actually spoken". She promoted a new philosophy of language which, because it derived both from *the Tractatus* and *the Investigations*, embodied fundamental and original logico-scientific insights about language which could be tested out on computers. The philosophical insights which provoked this practical application were, first, that a language is primarily a totality of contexts, not of sentences or words, and secondly, that contexts can only be distinguished from one another by using analogy.

These insights contributed to Masterman's context-based view of language and led to the idea of a thesauric interlingua method for MT. She borrowed from Wittgenstein's *Investigations* that a concept in language is like a Gestalt figure (one can "see" it differently by looking at it under various aspects) and that each of these "aspects" can be represented as a context. The different contexts of a word, undefinable and indistinguishable by normal methods, can be distinguished from one another by giving the sentence in which each occurs an analogy. Some or all of what Wittgenstein said about rules, logical grammar, notations and language games, she said, can be found even by looking at perfectly ordinary language. Among the new philosophers' insights adopted by Masterman and the CLRU, were J[ohn] L[angshaw] Austin's (1911-1960). In " How to talk : some simple ways " *Proceedings of the Aristotelian Society*, June 1953) Austin claimed that the primary patterns of how we actually talk (and think) are simpler and more fundamental than, and different from, grammatical pattern ; following Austin again ("a plea for excuse" *Presidential address to the Aristotelician Society Proceedings* October 1956), Masterman assumed that dictionaries could be processed to produce closed circles of semantically analogous, rather than synonymous, definitions. In other words they could be used to generate thesaurus-like headwords.

It must be added that, in order to process a thesauric interlingua method of MT, formalization was needed. Contexts were formalized by using lattices which, contrary to simple trees, allow inheritance of the concept properties from multiple supertypes. Thus the different contexts of a sentence were viewed as a sub-lattice of a thesaurus. However, for the CLRU, formalization did not come first. Language has to be considered as a whole, and mathematically formalizable only as a second step, whereas

for Bar-Hillel it is the opposite : language is considered as mathematically formalizable *a priori*; and it is the researcher's task to discover how natural languages can be adapted to formalization.

Let us note, as a conclusion, that in spite of their elaborated answer, the CLRU did not succeed in convincing Bar-Hillel, and they suffered, as the other MT groups, from the shortage of funds which resulted from the ALPAC report. However, until Masterman's death in 1986, they continued with diversified projects in Natural Language Processing, such as information retrieval, artificial intelligence and machine-readable dictionaries.

REFERENCES
Bar-Hillel, Yehoshua. 1953. A Quasi-Arithmetical Notation for Syntactic Description. In: *Language* **29**, 47-58.
Bar-Hillel, Yehoshua. 1954. Logical syntax and semantics. In: *Language* **30**, 230-237.
Bar-Hillel, Yehoshua. 1960. The present status of Automatic Translation of Language ", In: Franz L. Alt (ed.), *Advances in Computers,* vol.1, New York and London : Academic Press, 91-141.
Chomsky, Noam. 1955. Logical Syntax and Semantics. Their linguistic relevance. In : *Language* **31**, 36-45.
Language and Machines. Computers in translation and linguistics. 1966. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council.
Léon, Jacqueline. 2000. Traduction automatique et formalisation du langage. Les tentatives du Cambridge Language Research Unit (1955-1960). In : Piet Desmet, L.Jooken, P.Schmitter and P.Swiggers (eds.) *The History of Linguistics and Grammatical Praxis.* Louvain and Paris : Peeters, 369-394.
Léon, Jacqueline. 2002. From universal languages to intermediary languages in Machine Translation : the works of the Cambridge Language Research Unit (1955-1970). In : *ICHoLS 9, 9th International Conference on the History of Language Sciences, Sao Paulo (Brazil), August 27-30 2002.*
Masterman, Margaret, A.F. Parker-Rhodes, K Sparck Jones, Martin Kay, E.B. May, R.M. Needham, E.W. Bastin, C. Wordley, F.H. Ellis, R. McKinnon Wood, *Essays on and in Machine Translation by the Cambridge Language Research Unit,* dedicated to Yehoshua Bar-Hillel, June 1959, unpublished.
Masterman, Margaret. 1959a. Fictitious sentences in language. In: *Essays on and in Machine Translation by the Cambridge Language Research Unit.* ML91. unpublished
Winograd, Terry. 1972. *Understanding Natural Language.* New York : Academic Press.
Wordley, C. 1959. Establishing the contextual meaning of words. In: *Essays on and in Machine Translation by the Cambridge Language Research Unit.* unpublished, ML91, *Appendix II.*