[From: *BISFAI'95: Proceedings [of] The Fourth Bar-Ilan Symposium on Foundations of Artificial Intelligence*, ed. Moshe Koppel & Eli Shamir, June 20-22, 1995, Ramat Gat and Jerusalem, Israel (Menlo Park, CA: AAAI Press, 1996)]

Bar Hillel and Machine Translation: Then and Now

Sergei Nirenburg
Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003, USA
sergei@crl.nmsu.edu

The name of Yehoshua Bar Hillel is known to every student of machine translation. The late philosopher richly deserves this recognition: he had been a central figure in the early development of the field and contributed what should be considered the first set of sober assessments for MT at the time of the widespread gung-ho attitude to its prospects. The fame is probably due to the correctness of Bar Hillel's forecasts of the field's future stumbling blocks. Still, what is generally remembered of Bar Hillel's contributions to MT is but a small fraction of his ideas and opinions on the subject.

There is an unsettling tendency among my colleagues in MT, natural language processing, linguistics and AI tacitly to assume that their predecessors in the field, not having at their disposal either the latest machines or the latest theories, were somehow naive and "incomplete." History of the subject starts for them with the dissertation work of their thesis advisor. For such people, Bar Hillel may only have the distinction of being the earliest widely quoted author in MT. In reality, reading Bar Hillel can be very instructive for today's scholars, especially as he excelled in recognizing and assessing the intellectual evolution of entire fields (philosophy, logic, linguistics, MT) and pointing out lacunae in the applicability of their findings.

This paper, first, takes the reader on a brief tour of Bar Hillel's writings about MT and, second, attempts an assessment of the current state of MT research in the spirit of Bar Hillel (though without any claims for similar profundity).

1. Then

Bar Hillel's central argument can be presented as the following sequence of brief theses. The summaries are illustrated by a few quotations.

1. Bar Hillel believed that, in order to achieve fully automatic high-quality machine translation (FAHQMT), machines must be able to process meaning. "The task of instructing a machine how to translate from one language it does not and will not understand into another language it does not and will not understand presents a real challenge for structural linguists, in that their thesis that language can be exhaustively described in non-referential terms undergoes here an experimentum crucis. If, in a translation programme, some step has to be taken which directly or indirectly depends on the machine's ability to understand the text on which it operates, then the machine will simply be unable to make the step, and the whole operation will come to a full stop. (I have in mind present day machines that do not possess a semantic organ. The situation might change in the not too distant future.)" Some Linguistic Problems Connected With Machine Translation. *Aspects of Language,* p.308. In this passage, Bar Hillel addresses

translation without treatment of meaning. "Non-referential" means "uninterpreted" that is, free of meaning. “It now seems that for the purpose of computer-aided translation the semantic structure of sentences to be translated has to be exhibited.” The Outlook for Computational Semantics. *Aspects of Language,* p.358.

2. The study of meaning in language is the realm of semantics, and there were at the time two major traditions that could be drawn upon to provide solutions for MT. These were linguistic semantics and logical semantics.

3. Contemporary linguistic semantics (epitomized by the work of Katz and Fodor, 1963) was found largely unpromising by Bar Hillel because a) it was not adequately formalized; and b) its apparatus neglected large parts of semantics such as non-compositional phenomena, and every other phenomenon for whose description it is necessary to rely overtly on “background information,” the knowledge of the world. Note that Bar Hillel actually saw something positive in Katz and Fodor’s program, in that he believed that linguistic semantics can overcome its shortcomings by paying attention to issues not developed by Katz and Fodor: “It is very likely that their shortcomings will further the field much more than their actual positive achievements.” Review of “The Structure of Language.” *Aspects of Language,* p. 175

4. Contemporary logical semantics was dismissed by Bar Hillel in the context of MT because it focused on artificial languages. “... [Rudolf Carnap] is thinking mostly in terms of constructed languages...” The Outlook for Computational Semantics, *Aspects of Language,* p.359.

5. Bar Hillel believed that treatment of meaning can only be based on a system of logic: first, because for him only hypotheses formulated as logical theories had any scientific status and, second, because inference rules necessary for MT could only be based on logic.

6. At the same time, he considered such logical systems *unattainable* because they could not work directly on natural language, using instead one of a number of artificial logical notations. “...The evaluation of arguments presented in a natural language should have been one of the major worries... of logic since its beginnings. However, ... the actual development of formal logic took a different course. It seems that ... the almost general attitude of all formal logicians was to regard such an evaluation process as a two-stage affair. In the first stage, the original language formulation had to be rephrased, without loss, in a normalized idiom, while in the second stage, these normalized formulations would be put through the grindstone of the formal logic evaluator. ... Without substantial progress in the first stage even the incredible progress made by mathematical logic in our time will not help us much in solving our total problem.” Argumentation in Natural Language, *Aspects of Language,* pp. 202-203. Bar Hillel criticized the methodology of logical semanticists because they spent all their time on a partial task, without concern for complete coverage of language phenomena. “One major prejudice... is the tendency to assign truth values to indicative sentences in natural languages and to look at those cases where such a procedure seems to be somehow wrong...” Argumentation in Natural Language, *Aspects of Language,* p. 203. There were in Bar Hillel’s time no extant proposals about how to translate natural language into a formal language amenable to processing by logic. The reason, according to Bar Hillel, was that such a translation process would have to rely on knowledge about the world.

7. Thus, acquiring world knowledge became a precondition for the success of an entire sequence of enterprises culminating with machine translation. Bar Hillel considered this task infeasible, and this was the ultimate reason for his well-known pessimism about MT and computational semantics. "It seems now quite certain ... that with all the progress made in hardware, programming techniques and linguistic insight, the quality of fully autonomous mechanical translation, even when restricted to scientific or technological material, will never approach that of qualified human translators and that therefore MT will only under very exceptional circumstances be able to compete with human translation. This "pessimistic" evaluation is based upon various considerations, only one of which will be presented here, and even this, for obvious reasons, only very shortly and therefore dogmatically. Expert human translators use their background knowledge, mostly subconsciously, in order to resolve syntactical and semantical ambiguities which machines will have either to leave unresolved or resolve by some "mechanical" rule which will every so often result in a wrong translation." The Future of Machine Translation, *Language and Information,* p. 182.

Bar Hillel's famous example concerned the following text: *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.* "Why is it that a machine ... is ... powerless to determine the meaning of pen in our sample sentence within the given paragraph? The explanation is extremely simple, and it is nothing short of amazing that, to my knowledge, this point has never been made before, in the context of MT... What makes an intelligent human reader grasp this meaning so unhesitatingly is ... his knowledge [of] the relative sizes of pens, in the sense of writing implements, toy boxes and pens in the sense of playpens... Whenever I offered this argument to one of my colleagues working on MT, his first reaction was: "But why not envisage a system which will put this knowledge at the disposal of the translation machine?" Understandable as this reaction is, it is very easy to show its futility. What such a suggestion amounts to, if taken seriously, is the requirement that a translation machine should not only be supplied with a dictionary but also with a universal encyclopedia. This is surely utterly chimerical and hardly deserves any further discussion." Nonfeasibility of FAHQMT, *Language and Information,* p. 176.

8. The above line of reasoning led Bar Hillel to the conclusion that FAHQMT should not be the stated goal of MT researchers, as significant practical advances could be obtained for less demanding objectives. "...[T]here is really no need at all to compromise in the direction of reducing the reliability of the machine output. True enough, a smooth machine translation looks impressive... [but] it is much safer to compromise in the other direction. Let us be satisfied with a machine output which will every so often be neither unique nor smooth, which every so often will present the post-editor with a multiplicity of renderings among which he will have to take his choice, or with a text which, if it is unique, will not be grammatical. ... Let the machine ... provide the post-editor with all possible help, present him with as many possible renderings as he can digest without becoming confused by the embarrass de richesse ... but never let the machine make decisions by itself on purely frequential reasons even if these frequencies can be relied upon." Aims and Methods of MT. *Language and Information,* p. 171.

2. Now

Our MT research group at CRL shares all of Bar Hillel's premises, including that concerning unattainability of a complete database of world knowledge. We believe, however, that world models are useful and feasible even when they are not complete or provably correct. This seems to be in the spirit of Bar Hillel's own opinion concerning competence and performance theories: "I have already voiced ... my misgivings over the conception of Chomsky and others that a more or less complete development of a theory of competence is a prerequisite for the development of a theory of performance. That this could not be so can be seen simply from the fact that the very adequacy of a particular theory of competence can only be determined on the basis of performance, with or without theory" Review of John Lyons' "Introduction to Theoretical Linguistics". *Aspects of Language*, pp371-72.

Our approach to FAHQMT-oriented research (e.g., Nirenburg et al., 1992, Onyshkevych and Nirenburg, 1995) is based on the centrality of meaning and on a logical system underlying meaning. The logical system is interpreted with the help of an ontology, a world knowledge model.

However, at present, MT research and development at CRL and elsewhere, incorporates work whose objectives cover the entire spectrum between theoretical and empirical linguistic approaches, fully-automatic MT and machine-aided human translation tools and, most prominently, between "rule-based" and "corpus-based" approaches, where the latter label applies to systems which treat language as an artifact (the sum of all texts written in it or at least all texts available in electronic form) and apply general-purpose statistics-based methods to solve problems such as translation. The debate between the rule-based and corpus-based approaches has enlivened the field of computational linguistics since about 1990. One possible assessment of the state of affairs in this debate follows.

Machine translation has been a fashionable field for at least forty years of its fifty-year history. The reasons for this vary from R&D glory to commercial payoff. Over the years, an impressive variety of methods have been used as the basis for translation programs. The problem has, however, proved so complex that the quality of the final result has not correlated significantly with the method chosen. Rather, it correlated with the amount of descriptive work on language that was carried out.

Of course, MT research has brought about significant side benefits. Entire scientific fields were created largely due to MT efforts: witness the nascence of computational linguistics. Often, MT was used as an application of choice for a variety of workers to test and attempt to corroborate their theories of language and of human thinking capacity. It is characteristic that the final report of the Eurotra project listed as its major success the creation of computational-linguistic infrastructure in the countries of the European Community deemphasizing the fact that no realistic MT system was built under its auspices. Many factors contributed to the lack of the engineering achievement in this project, among them the relative lack of accent in Eurotra on actual description and system building, with preference given to designing detailed formal specifications of (largely syntactic) levels of analysis and their corresponding formalisms.

Is the Eurotra case prototypical for the entire field of MT? One of the problems with the field has been that the descriptive work is, frankly, rather monotonous and boring. This is why attempts were made either to make it less boring (by adding an independently motivated theoretical angle to the descriptive work) or to try to avoid it altogether.

The latter objective was made manifest in a) attempts to use AI learning techniques or more practical semi-automatic procedures for knowledge acquisition and b) the application of statistical methods for establishing cross-linguistic correspondences in lieu of language description work. The former solution made itself manifest in viewing MT as a testbed for one's favorite linguistic or computational-linguistic theories, such as the currently fashionable "principle-based" approach to syntax. Machine translation is indeed a tempting avenue of computational inquiry into modeling human mental and language processes, and a number of approaches to NLP in AI dabbled in MT as a potential application. Knowledge-based MT is a direct offshoot of the AI tradition.

The most remarkable feature of the statistical methods in MT is that they are not at all specific to their subject matter — the same techniques applied to processing language could and are used, for example, in the studies of the human genome.

The current R&D-oriented MT approaches, whether rule-based or statistical or hybrid, are based on "imported" ideas. At the same time, the best systems on the market cannot boast much by way of technological or scientific advances. Instead, they rely on brawn: huge, handcrafted dictionaries and grammars and a plethora of specialized translation routines. All of us are curious to see how well the R&D approaches will work once sufficient resources are allocated for one or more of them to reach the status of a product. The question is: what kind of imported techniques shows the most promise? The answer is not clearly obvious and is determined by sociological (read: the vagaries of funding) as well as scientific and technological trends.

The major scientific (or methodological) trend in the field is experimenting with how well the statistics-oriented methods will advance the state of the art in MT without the need for massive manual knowledge acquisition.

The major technological trend in the field is looking for the best ways of mixing the statistical and the "rule-based" methods. This author has been an early advocate of mixing such methods at the level of their final results, a method called multi-engine MT. Other approaches seek a more involved interaction, with statistics used not only during the process of MT but also to support development of background resources (i.e., dictionaries and grammars).

The major sociological trend, at least in the US, is the emphasis on a regimen of evaluations and competitions among MT (and, more broadly, NLP) systems. This promotes rigor and discipline as well as conformity and search for local solutions, which are not necessarily the most promising ones in the long run. Approaches that show a steady improvement are rewarded. Approaches with long gestation periods are punished.

Emphasis on mixed approaches is, for non-statisticians, a rearguard regrouping action, while for statisticians (witness the evolution of the claims and practices of the Candide IBM MT group), a

search for any avenue for improving the rather modest final results.

The knowledge-based and linguistics-based methods will do good by regrouping and concentrating on those tasks and situations in which statistical approaches fail to deliver. One must, however, remember the lesson of computer chess: at present, the best chess-playing systems are not terribly knowledgeable about chess strategy and tactics but they consistently beat AI-based programs and compete on equal terms with grandmasters. The $64,000 question is: how much more complex is human translation ability compared to the human chess-playing ability? That is, for how long will there be an opportunity to study language use through MT? If statistical methods succeed, rule-based MT may go the way of the AI-based chess programs.

MT seems to be too complex a task to be fully accountable for by the current statistical processing methods, even though these methods do not aspire to building representational models of human language capacity and rely only on the input-output behavior of such models (in MT, a text and its translation). In the final analysis, the open-endedness of language will become the stumbling block for these methods conceptually, just as logistically, the chronic shortage of resources (bilingual corpora) may precipitate the swing of the pendulum of MT R&D fashion back to the mentalist camp from its current behaviorist direction.

How long will this take? If history is any guide, such swings come roughly every 30 years: mentalism was in scientific ascendancy between 1960 and 1990, while behaviorism reigned, at least in the US, for about thirty years prior to that. Of course, we cannot be certain that we are witnessing this pendulum swing and not some other, unconnected development. Time will show. A more intriguing thought is that, just possibly, the rule-based/corpus-based dichotomy is not as important as we currently think. Maybe the real problem of MT as technology is that it is not generally understood how difficult the problem actually is. The confident claims, made by newcomers to MT (including this author some fifteen years ago), help stoke the high expectations of getting the desired result with a modest expenditure. At the current level of MT R&D, either the expectations should be lowered or the time scale of getting the results must be significantly extended.

The above list of opinions, though quite current in 1996, would not have been entirely out of place around 1960, the time at which Bar Hillel's celebrated assessments of MT were published. I think the entire enterprise of MT will continue to be reasonably successful (recurring outlandish claims coming from confident newcomers notwithstanding) as long as at least some of Bar Hillel's legacy of self-assessment is preserved.

References

Bar Hillel, Y. 1964. **Language and Information.** Reading, MA: Addison Wesley.
Bar Hillel, Y. 1970. **Aspects of Language.** Jerusalem: Magnes.
Katz, J. and J. Fodor. 1963. The Structure of a Semantic Theory. *Language,* 39, pp. 170-210.
Nirenburg, S., J.Carbonell, M.Tomita and K.Goodman. 1992. **Machine Translation: A Knowledge-Based Approach.** San Mateo, CA: Morgan Kaufmann.
Onyshkevych, B. and S. Nirenburg. 1995. A Lexicon for Knowledge-Based MT. *Machine Translation,* 10, pp. 5-57.