3 The Wall

3.1 The fascination of universal sememes

As mentioned in the previous chapter, I was part of a machine translation project that began in 1970. We were part of the linguistic camp that believed in the importance of using a comprehensive model of syntax, and our model was Junction Grammar, developed by Eldon Lytle, the project director. In fact, the project was at least in part a test of Junction Grammar. As we saw in the previous chapter, by 1960 people like Bar-Hillel were claiming that there is a wall between human translation and machine translation; and in terms of performance on unrestricted text, there is now general agreement that such a wall exists. But what kind of wall is it? Where does it come from? Can it be taken down? Throughout the rest of the book, the wall metaphor will evolve as we discuss language. I will begin by discussing my personal encounter with the wall during the 1970s. The wall has turned out, to my surprise, to be more durable than the Berlin Wall.

When our machine translation project began, I was blissfully ignorant of any wall between human and machine translation. We were still using punched cards on an IBM mainframe, and I was focused on the hill we carried our cards up each morning to get to the computer center. Tired of returning in the evening to see the results of our daily computer run, we soon discovered an alternative, the VM/CMS operating system, which allowed us to interact directly with the mainframe using video display terminals. Not only did this allow us to be much more efficient in programming, it also opened up the possibility of on-line interaction with computer programs, which is familiar to everyone today who uses a personal computer, as opposed to the previous **batch** mode. We were translating from English into several target languages, so the cost of analyzing the English source text could be shared over several target languages. By then we had already bumped up against the wall, but we did not recognize its importance and hoped that on-line interaction would get us over the difficulties.

The interactive process of analysis consisted in running a computer program that attempted to analyze the source text but paused whenever an ambiguity was detected, displaying a question on the computer terminal and waiting for a human to answer it. The human answering the questions was an English monolingual with some training in Junction Grammar; therefore, no questions requiring any knowledge of the target language were permitted. The questions were of two types, syntactic and word-sense.

Syntactic questions involved aspects of Junction Grammar. Some ambiguities were easy to state and to answer, such as where a prepositional phrase attaches in a sentence (e.g., in the sentence "I saw the girl with a telescope" the prepositional phrase can attach to 'girl' or to 'saw', depending on whether the girl has a telescope or someone used a telescope to see the girl). Other interactions were more subtle, such as the possibility of attaching a modifier to several different points within a noun phrase or verb phrase, with point of attachment having semantic implications.

Word-sense interactions were of an entirely different nature. Here Junction Grammar gave us no clear guidance. Neither did we have much guidance from other machine translation projects. All machine translation projects, from the early 1950s on, recognized the importance of resolving categorial ambiguity. As indicated in the previous chapter, this process has traditionally been called homograph resolution in machine translation circles. For example, the LTU 'open' can function as a verb or an adjective ("We will open the case of oil" versus "We will keep the open case"). Clearly, resolving ambiguities of grammatical category is necessary in machine translation, but just as clearly, this simple categorial resolution is not always sufficient to determine the proper target language word. For example, the word 'beam' has several distinct noun senses, such as "beam of light" and "a beam of wood". It is not likely that these senses will translate as the same word in all target languages; thus, categorial resolution is insufficient, and the computer cannot simply look at the next few words since an ambiguous word such as 'beam' will often be used without a prepositional phrase. When a word with multiple senses was encountered, the machine translation system attempted to resolve the categorial ambiguity automatically, but when a word had multiple senses within a category, the program presented a list of definitions to the English monolingual to choose from and waited for a reply.

Other machine translation projects had of course encountered these same problems, but the other projects were different in design and objectives from ours. They were usually translating technical documents. In order to increase the chances of the computer making the right choice, the designers allowed the user to select among several technical dictionaries when a text was translated, with each dictionary tailored to a particular subject matter domain, such as chemical engineering or nuclear physics. Within each technical dictionary, a term was given just one translation, the one most likely to apply to documents specific to the domain in question. In case a document dealt with several domains, the user could even specify that several technical dictionaries should be consulted in a particular order. As we saw in the previous chapter, the Systran system had only one general translation for each word, but it was only used when the word was not found in any of the technical dictionaries specified for use during the translation of that text. This method was surprisingly effective when applied to documents within a very narrow domain. In contrast, we were translating rather general documents, often ranging over several topics within a single text, so we did not have the luxury of using a domain-specific approach. Also, most other projects were committed to a design in which the machine translation phase was automatic and produced translations that were as good as was feasible under that constraint, with the hope that their quality would gradually improve and the assumption that post-editing would be used as needed. Again in contrast, we were committed to producing high-quality output, even if this required a large number of interactions at first, with the hope that the number of interactions per sentence would gradually be reduced without sacrificing quality.

At first we thought we were alone in using the interactive approach, but soon we ran across an article by Martin Kay (1973) which reported on his implementation of interactive translation in a small system. With this and other encouragement, we pursued our attempt to develop a large-scale system for producing interactive high-quality machine translation of unrestricted text. Little did we know what awaited us.

Preparing a dictionary for interactively resolving word-sense ambiguities seemed at first to be a well-defined task, at least as well defined as the senses listed in a good dictionary. But questions quickly arose. For a given word and a given grammatical category, which word senses should be listed as options? Should they be tailored to the various translation equivalents in the target language? Tailoring was not acceptable since there were several target languages in our system, with the possibility of adding more at any time. Here I must mention an additional design requirement based on economics. Since we anticipated that interaction during analysis would be expensive, we wanted to be able to do it once for a given text and then re-use the results of analysis, sometimes much later, to translate automatically into several target languages, including target languages that were not yet part of the system. Therefore, the obvious approach was to identify all the possible concepts onto which a given word can be mapped. In order to build analysis dictionary entries that would not have to be redone (except to add new concepts as the language evolves) and to avoid having to re-analyze an old text even in that case, we decided to use only language-independent concepts and to assign a unique number to each of them. Then, we could map each universal, language-independent concept, which we called a **sememe**, onto just one word (or a small cluster of grammatically conditioned words) in any target language we chose to translate into.

Although the search for language-independent concept numbers now sounds to me misguided, twenty years ago we were extremely serious about the endeavor and convinced of its eventual success. So we embarked on more than five years of concerted effort to write an interactive Junction Grammar analyzer, to build up an analysis dictionary that incorporated many word senses which we hoped came from the language-independent universal set of sememes, and to write transfer/synthesis programs into several languages. Although we believed that the syntactic relations in Junction Grammar were rich enough to represent the structure of any sentence in any language, we did not pretend that our syntactic diagrams were language independent. Therefore, we did not have a true **interlingua**, that is, a language-independent intermediate representation, yet we did not find it strange to expect our sememe numbers to be language independent.

I remember a conversation with Richard Kittredge during this period (the mid-1970s). Kittredge was the director of the machine translation project at the University of Montréal, the TAUM project, and he suggested that our project was overly ambitious. I assured him that thanks to the power of Junction Grammar and the technique of using human interaction for the moment (with the goal of gradually reducing it to a negligible level or better yet, eventually eliminating it through intelligent programming), we were guaranteed to achieve success. I am now embarrassed when I recall this conversation.

Why did it seem so natural to me and to others who were working on the BYU project to attempt to write definitions which corresponded directly to both word senses and universal sememes? Recall from chapter 1 that the Western belief in a stable transcendental meaning, which is the basis for universal sememes, can be traced back many centuries, and that the shift away from stable transcendental meaning began only in this century with Saussure. Since Saussure and his contemporaries we have been in a confusing period of abandoning old ideas while not being satisfied with new ones, and often holding contradictory views simultaneously. Perhaps I can explain my willingness to believe in universal sememes by noting that Descartes, a great mathematician as well as a philosopher, found it worthwhile to try to do exactly what we were attempting to do with our dictionary of correspondences between words and sememes. Karen Sparck Jones (1986:211) cites Descartes who, in a letter to Pierre Mersenne dated November 20, 1629, described a proposed universal language built up from a basic set of "clear and simple ideas". Each basic idea would receive a unique numeric code; texts written in these codes would supposedly be understandable to anyone possessing a code book for the language. Later, in a well-known effort, Leibniz continued the search for a universal language. If minds as great as Descartes and Leibniz believed in what we were attempting, why should I be ashamed of our efforts?

3.2 Hitting the wall surrounding domain-specific machine translation

The syntactic side of our project went well. Junction Grammar developed into a coherent, fairly comprehensive model of the syntactic structure of language (Lytle 1974 and 1980; Melby 1980 and 1985), and many Junction Grammar insights have since occurred to others and have been incorporated into mainstream syntactic theory. The word-sense side was a disaster. As the dictionary got bigger and bigger and we ran more and more texts through the analyzer, it became harder and harder for native speakers of English to decide among the word-sense options listed for a given word. Often, they would say that a word was used with a sense somewhere between two listed senses but not right on either one. Often, they would say that an entirely new sense was needed, not like any on the list. And often, even if they chose a word sense from the list, the transfer/synthesis program would not produce an appropriate target-language word. The transfer/synthesis designers would complain that the analysis people were not choosing word senses consistently; the analysis people would complain that the dictionary writers were not doing their job; and the dictionary writers would explain that they were doing their best. It was a frustrating time, because no one could come up with a single example of a sentence that could not be handled perfectly—in hindsight, that is. Let me explain. Given a sentence that did not translate correctly the first time through the machine translation system, a diagnosis was made to identify the nature of the problem. Sometimes the sentence had been assigned an incorrect syntactic structure, and the analysis grammar team fixed the problem. Sometimes a sememe was mapped onto the wrong word in the target language, and the transfer/synthesis team fixed the problem. Sometimes there was a word in the

English source sentence for which there was no definition in the analysis dictionary that corresponded exactly to the meaning of the word in that sentence of the source text, and the analysis dictionary team fixed the problem. Every problem could be fixed, but the system usually failed on the next sentence presented to it. We did not know it at the time, but it was a form of the tuning problem that has been encountered by most, if not all, machine translation projects. There is no problem that cannot be fixed, yet you are never sure you have finally got it right.

By 1978, we had reached a crisis. We had hit a wall we could not get over. The system was not working (that is, the unfortunate need for heavy post-editing was recognized) and drastic action was in order. Some favored adding interaction to the transfer phase. But this was not an acceptable solution to me, since it required a bilingual human (actually, a qualified translator) for each target language (creating an economic problem) and since it did not answer the question of how to identify language-independent universal sememes (sidestepping a theoretical problem). This was when it suddenly occurred to me that although concepts certainly exist in some way, the real problem could be that the *language-independent universal sememes* we were looking for do not exist! Seldom in a person's career does such a complete change in world view take place. I had a degree in mathematics. The world was supposed to be a nice, tidy place. It was hard for me to admit that our approach to word senses was dead wrong. It was not just a matter of trying harder. I eventually became convinced that no amount of effort would patch up the system as we had designed it. The universal sememe search had reached a dead end.

My view of language and the world had changed. I had become convinced that what we had been looking for, *the* universal, language-independent set of sememes, did not exist. I started asking myself about the status of sememes. What was the justification for believing in them? It seemed to be an unsupported generalization: just because some terms could be reduced to a limited set of clearly distinguishable language-independent senses, we had supposed that all words could be. But pushed to the extreme, this generalization entails the doctrine of total reducibility of meaning to language-independent atoms, which we had called sememes. When we map all language, not just a few terms, onto sememes, supposing that there are a limited number of sememes that each word can map onto, we leave true creativity behind. All meaning becomes mechanical combinations of atomic word senses. Once this work is completed, we would *theoretically* arrive at the single correct, language-independent, all-encompassing basis for viewing, that is, categorizing the world, one universal view from which all individual views would be derived. This position has implications which, upon later reflection, I have found horrifying. In the early 1970s, I scoffed at the Whorf-Sapir hypothesis that one's world view is influenced by one's language. Now I have read some of Whorf's writing with interest and respect.

I emphasize that at the time, my objection to the view I abandoned included no high-flying reasoning. What brought down my naive faith was an *intuitive* realization that the search for the final determinate set of sememes was futile, obviously in practice but, I suspected, also in principle. I was beginning to sense that language is more dynamic and ambiguous in a deeper way than I had ever supposed and that there is not just one built-in, correct way to view the world. With no universal world view and thus no universal set of sememes, the basis for all lexical transfer in our machine translation system had disappeared.

The futility of searching for universal, language-independent sememes could explain the failure of *our* project. However, certain other machine translation projects had been successful. Why were they successful? I had already noted differences in approach. Upon comparing their design and objectives with ours, I observed the following that our project had the following design requirements: (1) it must produce high-quality translations (indistinguishable from competent human translations) of general texts; (2) human interaction in analysis was allowed, but little or no post-editing was permissible; and (3) all interaction with the computer must assume an English monolingual human operator. It turned out that no other project anywhere in the world was at that time able to satisfy those requirements; therefore, successful projects must have other design requirements. The crucial difference in design between our project and other projects was that we were not able to restrict the source text to the sublanguage of a single, well-defined domain. There was some kind of wall between sublanguage machine translation and general-language machine translation, but I did not at the time understand the nature of that wall.

3.3 The nature of the wall: two kinds of language

By 1981, the machine translation project had moved off campus, I had decided to stay on campus, and I had begun studying human translation. I was suddenly fascinated by human translation in its own right, since skilled human translators were doing, and had been doing for thousands of years, something which no machine translation project had been able to do—produce high-quality translations of a variety of general-language texts. I decided to try an inside-out approach. I became a translator, at least in the sense that I passed the accreditation exam (for French into English) and became a member of the American Translators Association. I even did some translation of articles on linguistic theory and contracted with a company to supervise the translation of a software manual into French. I also became very much engaged in the task of designing and programming productivity tools for human translators. The last few years of the BYU project, in addition to working on interactive machine translation (but of a basically different design once interaction in transfer was added in 1978), we also worked on various computer-based tools for translators. The development of these tools, along with interactive machine translation, was pursued off campus at ALPS (later ALPNET) by a group whose core consisted of former members of the BYU project. Another thread of development was pursued on campus and off campus (at LinguaTech) by myself and others. Once again, we ran into an insightful paper by Martin Kay (1980) that showed he was thinking along the same lines we were, and further.

Having been deeply committed to a certain view of machine translation and having been humbled by admitting that that view was in error but not having embraced a new view, I was prepared to see language in new ways. All through this period of the 1980s, in the midst of everyday activities of teaching, research, and software development, I yearned for a better understanding of what was wrong with the idea of a list of universal, language-independent sememes. Gradually, it dawned on me that within a well-defined domain—and I want to emphasize this restriction—one can get away with ignoring questions of universality. Each concept of a domain is normally associated with just one technical term per language. And in a standardized terminology file, each concept often has been assigned a unique number. This was exactly what we were trying to do for general language. So what was the difference? An answer struck me forcefully! *The concepts of a narrow domain and the concepts of general language are of a fundamentally different nature.* At first, this seemed too strange to be true. Technical terms and other specialized terms were thought to be just a subset of general vocabulary. However, on this new view, the general vocabulary of a language would form a language-specific semantic network, and each well-defined domain would have its own network, separate from the general-language network, which may be shared across various languages. In other words, there would be two fundamentally different kinds of network: general-language networks and domain-specific networks. This insight had overwhelmed me by 1989. If correct, it might explain much about when machine translation succeeds and when it fails. The next few years became a time to look for evidence for or against this general/domain distinction and to explore its consequences. The wall we had encountered between controlled-language MT and general-language MT turned out to be the wall around a domain that defines its boundaries.

3.4 The search for a term to cover both sides

Immediately, I realized that most texts consist of mixtures of general and domain-specific items. I say *item* because *word* does not quite fit technical terms, especially noun-noun compounds, that consist of more than one word. One piece of indirect evidence for a distinction between these two kinds of item came through the innocent search for a cover term for the two kinds. I have colleagues in the world of professional lexicography (makers of general dictionaries) and colleagues in the world of professional terminology research (translators and terminologists), and so I thought it would be a simple matter to talk with both groups to come up with a cover term that could refer to both expressions of general vocabulary (outside all walls) and terms from a specialized domain (inside a wall); that is, items on both sides of the walls around various domains. Not only was it not a simple matter, but in the unsuccessful process of trying to find a cover term acceptable to both, I discovered how different the two traditions are. Lexicographers use 'word', 'headword', 'compound', and 'expression', but terminologists do not. Terminologists use 'term', but lexicographers do not. The terms 'lexical unit' and 'lexeme' were proposed, but were rejected by the terminologists, who view their work as concept-oriented and the work of lexicographers as lexically oriented. I even considered using the neologism 'lexiconcept' suggested by a colleague.[8] I finally gave up and started using Lexical/Terminological Unit (LTU), as introduced in chapter 1. Unhappily, LTU is not a true cover term but rather a composite term. Looking for cover terms for similar items is not too difficult. A cover term for 'rose' and 'tulip' and similar plants is obviously 'flower'. But a cover term for 'rose' and 'hammer' that is not so general as to let in all kinds of other items ('object' would clearly be too general) might end up being a composite (such as 'flower/tool'). Thus, the lack of a cover term for lexical units of general vocabulary and terminological units of a specialized domain (other than a composite cover term such as LTU) suggests a substantial distinction.