

Chapter 9: Blending and NLP

9.0 Introduction

In this chapter, I will discuss research in computational Natural Language Processing (NLP), in light of the linguistic blending analysis in the previous chapters. The discussion and examples will focus on the problem of Machine Translation (MT) - the computational translation of text from one natural language to another. The aim of this chapter is not to propose computational models for natural language processing and translation. Rather, the aim is to analyze the theoretical implications of the blending analysis (in chapters 2-7) for future directions of research in NLP.

As Kay *et al.* (1994) note in the introduction to their book on the Verbmobil speech machine-translation project, “to many laymen, it is incomprehensible that we can build machinery that can convey a man to the moon, but none that can translate even very simple texts into French” (p.3). After many years of painful efforts in NLP, researchers still find it hard to identify what it is exactly about language and translation that defies all attempts of computational modeling, and just what could improve the quality of the result. In this chapter, I point to the role of linguistic blending operations as *one source* of difficulty for computational modeling of language processing and as an important direction for future research.

Maybe the most problematic aspect of computational language modeling is the influence of *context* on language interpretation. Numerous researchers in NLP (e.g., Kay *et al.*, 1994; Nirenburg *et al.*, 1992; Melby, 1995) have discussed the role of context as adding shades of meaning to linguistic forms, and influencing their translation. Input texts, it is suggested, cannot be processed in isolation: rather information from the contextual setting of the text must be taken into account. The term *context* includes both *linguistic context*

(i.e., the previous stretch of text or discourse and, how it influences the interpretation of the current utterance), *general world* (“common sense”) knowledge, and the *cultural* and *communicative* setting (see section 9.1.3). Kay *et al.* (1994) maintain that language is always *situated* in some contextual setting, and that the importance of the contextual setting is erroneously overlooked since speakers take it for granted. Discussing the role of the contextual setting in translation, Kay *et al.* note: “every professional translator is keenly aware that a great deal more than linguistic knowledge is required for the job That crucial knowledge that a translator must have has almost always been overlooked because it is shared by most humans, especially when they have a largely common culture” (p. 6). In recent years various attempts have been made to incorporate contextual knowledge into NLP systems. First, knowledge-based methods developed in AI (Artificial Intelligence) were adopted in NLP systems in the form of extra-linguistic world-knowledge data-bases that guide the linguistic processing of input texts (see section 9.2). Second, methods for computing textual context are also being developed¹ .

In this section, I will suggest that an additional crucial skill a translator (or any language speaker) possesses has been overlooked in research on computational modeling of language: the ability to creatively generate and interpret linguistic blends of the sort discussed in this dissertation (chapters 2-7). While contextual knowledge has played a central role in the examples studied in previous chapters (for example, in the *imposition* of prototypical cultural-experiential scenarios on interpretation of blends), the discussion in this chapter will concentrate not so much on the role of contextual knowledge in language

¹ While the various knowledge representation systems in AI have traditionally stressed representation of isolated *propositional* meaning, we do find recent attempts to incorporate considerations of the larger *textual context*. For example, Hovy (1988a,b) incorporates planning at the level of *multisentential* text in language generation. Nirenburg *et al.* (1992) report that in the computation of meaning of input texts, information about the “setting of the communication situation” is incorporated including parameters such as “the properties of the participants and their relative social status” (p. 75).

processing as much as on the necessary *mechanisms* for *manipulating* world knowledge and linking it to linguistic structures. For the discussion in this chapter, I will assume that "general world knowledge" (e.g., knowledge on prototypical scenarios in the world, or the functions of objects in the culture, etc.) is already coded in some form into the computational system. The issue I will focus on is how this knowledge is dynamically accessed and manipulated for the purpose of language processing. I will claim that the dynamics of mechanisms such as blending has been largely ignored in NLP, and that research on understanding and modeling such mechanisms must be pursued in addition to, or in parallel with, research on the coding (or statistical extraction) of "common-sense" knowledge. The discussion in this chapter focuses on the complex links that exist between *grammatical forms* and *semantic structures*. The sentences to be discussed are *isolated* (de-contextualized) sentences, so the role of discourse context (and the set of pre-assumptions it defines) will be ignored in this chapter.

The analysis in the previous chapters of the dissertation suggested that linguistic processing of even very basic clause structures (such as the English Caused-Motion construction, or the various Hebrew basic syntactic constructions and *binyanim*) involves complex mapping and integration operations. The language by itself provides only partial cues for the "de-integration" process: the "un-packing" and elaboration of linguistic utterances into conceptual representations. The analysis also suggests that much of the "integration" and "de-integration" operations are automatic, especially when triggered by very entrenched, conventional linguistic blends. What the entrenchment does is make the blending configuration less noticeable, but the blending schemas themselves (extracted and generalized from many linguistic instances) are still available for conscious processing. The conscious processing of blending operations is particularly noticeable in novel grammatical blends (e.g., in the coinage of novel root-*binyan* combinations in Hebrew - section 4.3, or

in novel lexical-syntactic combinations expressing caused-motion events in English - section 2.1), or during translation when switching between different blending conventions in different languages requires their conscious processing (chapter 8). But activation of blends also takes place in the everyday automatic elaboration of sentences during their interpretation (for example, the Hebrew stem *hif'il* marks that the event denoted by the root is part of a larger causal sequence. The *binyan* thus prompts the hearer to *elaborate* the semantic content of the linguistic structure, chapter 4). In this chapter, I will discuss the role of such "conscious" reconstruction of blending configurations in NLP systems.

Through the analysis in this chapter, I will make the following claims:

(1) Though many NLP systems incorporate vast amount of general world knowledge (in the form of hand-coded or statistically extracted rules), the use of these databases in generating a “functionally sufficient” semantic representation of linguistic structures is still very limited. In actual practice, contextual knowledge is used primarily to disambiguate the input text, but rarely to *add* information not explicitly provided in the text (information which may be necessary for further processing, e.g. translation).

(2) A mistake is often made when dealing with failures of an NLP system (e.g., failures in providing a correct “semantic model” or translation, for an input sentence) in assuming that the linguistic and world knowledge structures encoded in the system are necessarily inaccurate and should be modified or extended. Often the knowledge structures are accurate and complete, and the failure of the system results from the speaker’s *creative integration (blending)* of permanent knowledge structures into new temporary structures. The novelty of the input text is in the way the default structures are linguistically integrated together (and hence related to each other semantically in the sentence), and the goal of the system is to reconstruct these novel temporary blends for successful processing of the sentence. Modifying the permanent knowledge-bases of the system will not provide a general solution in such cases.

(3) Applying contextual world knowledge solely via "pragmatics" modules which *modify* an interpretation of a sentence *after* a basic semantic structure is computed is often ineffective. In the examples discussed in this dissertation, world knowledge *guides* the very basic assignment of a minimal semantic structure to a linguistic utterance.

(4) Pre-encoded inference rules can capture only the most entrenched (repeated) instances of blending. They cannot solve the core problem of blending. Reconstruction of blends has to be performed on-line, simulating human cognitive creativity in finding analogies and performing analogical mapping between retrieved knowledge structures and linguistic forms. To interpret a sentence in the view of this dissertation is to reconstruct a set of correspondences--a mapping--between a linguistic form and conceptual (knowledge) structures.

It should be noted here that though the discussion in this chapter suggests that current underlying assumptions of NLP research cannot generally support the findings of grammatical blending and translation in this dissertation, the discussion does not intend to imply that NLP research in its current form is ineffective. Not at all, because, in spite of the immense creative nature of thought and language (as reflected in the blending examples discussed in this dissertation), much of language use is in fact entrenched and predictable (if not in a deterministic way, then at least statistically). Current NLP systems can capture these repeated entrenched chunks of discourse and the conventional contexts when they are produced, and the partial success of NLP systems today shows that indeed these methods can produce acceptable results to some extent. In particular, the surprising relative success of statistical NLP (including statistical MT, see section 9.1.2.1) which is based on simply reiterating pieces of sentences from existing corpora, points to a prevalent trait of language generation by human: chunks of discourse are repeated by people over and over again. The analysis in the chapter claims however that for the future goal of highly automated NLP,

mechanisms for creative processing of language must be understood and incorporated as well.

The main challenge is the following: could automatic language processing (e.g., translation of even very simple technical texts) be done accurately enough without the incorporation of dynamic cognitive processes such as mapping, blending and integration of representational structures? My analysis suggests: probably not! And acknowledging the importance of such processes is the first step in enhancing NLP technology. Even if completely automatic modeling of analogical mapping and integration operations is not possible at this stage, some of the conceptual blending power can be incorporated into computational systems: for example, by encoding various levels of entrenched blends and using statistical information to choose among possible blends, or by incorporating human-machine interaction into NLP systems to interpret or generate grammatical blends (the latter issue will be discussed very briefly in the conclusions section 9.3). The incorporation of such mechanisms requires first that we *understand* them: we need to know *how* and *when* blending mechanisms take place in language processing in order to identify the kind of knowledge we need to encode in NLP systems, and how NLP systems could process this knowledge. In addition, even if blending mechanisms cannot be completely automated with current computational techniques, it is still important to realize which aspects of the *failure* of NLP technology are due simply to *scale* problem (which more powerful computers and better algorithms can solve), and which are due to the very *nature* of language processing versus current computational techniques² .

² It is often the case that the shortcomings of current NLP technology are attributed to insufficient advances in either of two underlying sciences:

(1) Formal Linguistics - The scientific understanding of the formal properties of grammatical systems is still far from being complete. A common assumption is that advances in the knowledge of formal linguistic systems will improve the performance of NLP systems.

The structure of the chapter is as follows: My analysis will focus on a sub-field of NLP, the field of Machine Translation (MT). I will start with a general presentation of the field of MT, its goals and its main approaches and strategies (sections 9.1.0-9.1.3). I will then go on to present what are typically considered to be the main problems in MT from the point of view of the system developers, and how these problems differ from the type of problems associated with translation of novel blends (section 9.1.4). In section (9.2) I will discuss the computation of semantic representation in NLP, and whether the techniques used today are capable of dealing with novel linguistic blends. Section 9.3 sums up the analysis in this chapter.

9.1. The field of Machine Translation - background

9.1.0 The prospects of Machine Translation

People interested in language and technology tend to react to the notion of Machine Translation (MT) in a passionate manner. Many are enthusiastic about the prospects of MT in the future. This is particularly evident in commercial circles as well as in science fiction literature. In the popular television series *Star Trek*, the computer in the star ship (Enterprise) can translate anything from any language. Visitors from other advanced planets have MTs installed in their heads. Fascination with MT is also shared among prominent figures in the computer industry. In a recent interview with Gordon Moore, the chairman of the board at *Intel* (*Yediot*, January 24, 1997). Moore predicted that the main advancements

(2) Artificial Intelligence and Knowledge Representation - NLP technology must rely on extensive databases of extra-linguistic general world-knowledge. However, there is currently no reliable and coherent way for representing general world knowledge computationally. Therefore it is often suggested that when the technology for representing knowledge in computers improves, so will the performance of NLP systems.

The analysis in this chapter suggests that even advancement in these two fields is not sufficient for high performance of NLP systems. Simulation of human cognitive skills (such as blending) is required in addition.

in computer technology at the beginning of the 21st century would be speech technology and MT. Moore believes that in the very near future we will be able to communicate with our computers using spoken everyday natural language, and converse on the phone in different languages with simultaneous computational translation. In contrast to enthusiastic followers of MT, we also find many who argue passionately that MT has no future: computers are so limited and translation is so complicated, that the whole idea of automated translation is impossible.

At the current stage of MT research and development, I believe both extreme approaches to MT are misguided. On the one hand, the goal of fully automated human quality translation is clearly far from our grasp, and the discussion in this chapter further suggests that present NLP techniques are not powerful enough to imitate the immense flexibility and creativity of human language processing. On the other hand, MT companies already provide customers with computational translation systems that perform economically profitable translation at various levels of automation and quality^{3,4}. The relative success of commercial MT systems today is still a far cry from the euphoria of the 1950s, when researchers seriously thought that the machine was going to take over the territory of translation as a whole⁵. What we see instead today is a redefinition of the original goals: instead of aiming at developing *fully-automatic high quality MT* (FAHQT),

³ The most prominent example of a successful MT system is SYSTRAN, a system developed based on work in the late 1950s and early 1960s at Georgetown University, which is still commercially active.

⁴ Experience with commercial applications of MT systems suggested that *low-quality* automated translation can be useful in certain contexts. Melby (1995:36) discusses the most prominent example: for the gathering needs of the U.S. airforce, scientists are expected to study relevant scientific articles written in Russian. Based on low-quality automated translation (MT), scientists can now select a small subset of the Russian articles for human translation.

⁵ None of the major basic research projects on MT in the world so far has attained the original goal of developing a high quality *fully automated* MT system (for example, both *Eurotra* - the major European effort in MT, and the Japanese *Fifth Generation Computer* project in which MT was a primary segment, ended without fully achieving their original goals).

current research efforts concentrate more on the goal of *Machine-Assisted Translation*, and the development of *translation tools* (Melby, 1995:41). From a scientific point of view, Machine Translation still remains one of the most intriguing domains for studying cognition and computation, and a primary test ground for linguistics models. Rather than dismissing the field scientifically (as some scholars do) because of its immense complexity, I believe that basic research should be directed at understanding where MT succeeds and where it fails in comparison with the human mind.

9.1.1 General strategies in MT: a brief history

During its early years, machine translation research was viewed as primarily an engineering task: translation was compared to a cryptographic code-breaking task⁶. The success of cryptography in breaking the Nazi code during World War II encouraged a view of MT as a feasible and attractive application of the new computer technology. Advances in linguistic theory and repeated failures of the first-generation systems to achieve their stated goals have united to discredit this attitude. Through the 1950s and into the following decade, machine translation has come to be understood as an application domain of formal linguistics and computer science, what would later become known as the discipline of

⁶ A known citation from Warren Weaver's (1955) original memo for MT best exemplifies the simplistic (engineering) view of translation and the underestimation of linguistic complexity. In Weaver's view, the linguistic content and structure of the translated text is exactly the same for all languages; only the encoding system differs: "When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode".

It is interesting to note however that many of Weaver's original suggestions have gained renewed popularity in a recent movement from rule-based MT towards statistical and corpus-based MT. Weaver noted that in contrast to cryptography, language involves ambiguity and therefore it is expected that a single word in the source text may have several possible translations. However, Weaver noted, if a human is allowed to see the word or two preceding and following the translated word, it is often possible to figure out what the word means and what its translation should be. This basic idea is used in statistical MT today which collects statistics on the translation of a word given its immediate context (preceding and following word in bigrams or trigrams). The basic problem of translation however still remains with this method: i.e., that translation of a sentence is not really a translation of its individual words, as will be discussed in this chapter.

computational linguistics.

The syntax-oriented approach of computational linguistics was criticized when it was demonstrated that fully-automated high-quality machine translation is possible only when some *meaning* of the input text is taken into account. The first to criticize contemporary MT research on this ground was Yehoshua Bar-Hillel (1959, 1960), who (looking back) focused on the role of *context* in *lexical disambiguation* of sentences. His now famous example was:

- (1) Little John was looking for his toy box. Finally he found it. The box was in the *pen*. John was very happy.

The word *pen* has (at least) two meanings - a writing tool and a playpen. Bar-Hillel's point was that a lot of practical information about boxes and pens, their use and their typical size is needed for deciding on the meaning and translation of the word *pen* in example 1. Bar-Hillel's criticism later led (in the 70's and 80's) to acknowledging that *general world knowledge* representation and manipulation is an important facet of machine translation, and Artificial Intelligence (AI) has been recognized as another field of which machine translation can be considered an application. In opposition to the Chomskyan generative linguistics view of the time that aimed at drawing a borderline between purely linguistic semantic knowledge and general world knowledge, research in AI assumed that there was no such line and that a semantic theory of language must include metalinguistic knowledge. The paradigm which strongly follows this line today in MT research is the one which has come to be known as *knowledge-based machine translation* - KBMT (e.g., Nirenburg *et al.*, 1992).

9.1.2 MT architectures

Traditionally, MT system architectures are divided between *Transfer*-based systems, and *Interlingua* systems:

In *transfer* systems, a source language sentence is first parsed into a syntactic (or syntactico-semantic) internal representation. Next, a transfer is made at both the lexical and syntactic levels into corresponding structures in the target language. In the third stage, a complete translation is generated. Two monolingual lexicons and one bilingual dictionary are needed in a Transfer system: a source-language and target-language lexicons, specifying basic syntactic and semantic attributes required for morphological analysis, parsing, and morpho-syntactic generation of the target language (e.g. part of speech, conjugation forms, basic semantic constraints), and a bilingual transfer dictionary (tailored for a specific source-target language pairs). In broad terms, the 'transfer' systems may be further divided into those based on *syntactic* transfer and those which go 'further' and incorporate lexical-*semantic* analysis to help resolve ambiguities in the source-language representation. In the latter system there is a continuous 'play' between the 'weight' given for the source text analysis (and disambiguation) and that given to the bilingual components. In many cases, lack of sufficient analysis and disambiguation in the source text analysis can be 'covered up' by a sophisticated bilingual transfer dictionary.

In *interlingual* systems, the source language and the target language are (theoretically) never in direct contact. Such a system has two clearly distinguished phases: first, source language analysis, which results in an artificial unambiguous formal representation of the sentence (such as that of frames or first-order logic). The second stage involves expressing the meaning (represented by the formal language) using the lexical units and syntactic constructions of the target language. Interlingua projects fall into two classes: the early syntactic approaches and those inspired by artificial intelligence. The former approach, which aimed at developing a universal syntactic structure based on Chomskyan theories of transformational grammar, was abandoned. The expressive power of the syntactic representation was found to be insufficient (Hutchins, 1986). The latter approach, which is based completely on AI techniques, makes use of inference mechanisms which apply

general world knowledge to the source text analysis and representation (see further discussion in section 9.2.1). The AI-based knowledge-based machine translation approach mentioned in the previous section belongs to this class of interlingual systems.

While transfer MT systems proved to be more practical and on average produced better results (e.g., the most successful system to-date is the transfer-based system SYSTRAN, the oldest commercially available system), interlingual MT systems provide several theoretical advantages over transfer systems. First, in translating from any one of n different languages to any of the remaining $(n-1)$ languages, $n(n-1)$ different transfer modules would ordinarily be required. But, if an intermediate language is implemented, transfer into each of the n target languages would use only the "universal" language (formal representation) as input; hence only n encoders from the source languages into the universal meaning representation and n decoders from the universal meaning representation into the target languages would be required in a transfer system, which would clearly be much more economical. The interlingua concept is also an important element in the modularization of the translation process. Modularization allows work to proceed independently on each sub-task, and it is usually easier to see the effect of changing or adding a rule in a modularized system (since the effects are localized to one language).

As Nirenburg *et al.* (1992:30) observe, the major distinction today between the interlingua- and transfer- based systems is not so much in the presence or absence of a bilingual dictionary (direct contact between source and target language) but rather in the attitude towards comprehensive analysis of meaning, or the *depth* of source language analysis. All rule-based MT systems involve a measure of linguistic analysis of the source language text. The purpose of the analysis is to facilitate the finding of target language correlates for the various meaning components expressed in the source language through its lexical units, syntactic constructions and word and sentence order. But while in "transfer" systems, the transfer of the source text into the target language proceeds directly from the

source text *syntactic* analysis stage, in interlingual systems some level of *semantic* representation of the source text is constructed from which the target text is generated. Therefore, the debate between interlingual and transfer approaches to MT today is in fact a debate on the role of meaning representation (in addition to linguistic syntactic representation) in translation. Interlingua approaches view meaning representation as a crucial step in the translation process.

It should be noted here that in using the term "interlingual" in defining MT systems, developers of knowledge-based MT systems today do not necessarily assume the existence of (or the possibility of defining) a complete "universal language" which can capture all communicated semantic content. Rather, the aim of knowledge-based "interlingual" MT systems is to generate "a *functionally complete* representation of meaning" - a semantic representation which is "(merely) sufficient for translation to a number of languages, rather than sufficient for total understanding, which entails a more complete, human-like inferential process for understanding all implicit and explicit information" (Nirenburg *et al.*, 1992:27).

The analysis of language processing in this study provides further support for the interlingual knowledge-based approach to MT by emphasizing the inescapable necessity of generating some form of independent representation of (partial) semantic content for a linguistic utterance before translation into the target language can proceed. The analysis in this dissertation particularly supports the goal of achieving only "functionally complete" representation of the input text semantics, since a basic assumption in the analysis is that no *one* single "meaning" exists for a linguistic structure but rather interpretation can potentially extend by association nets to theoretically infinite mental domains. For translation purposes, the "computation" of meaning representation of a source text needs to go only as far as required by the grammatical constructions, lexicon, and entrenched blending schemas of the *target* language.

The analysis in this chapter, however, differs from the conventional techniques practiced in interlingual MT systems today in that it emphasizes the fundamental need for *extending* the linguistic content of the source text before translation into the target language, rather than (as often practiced in MT systems) just representing the linguistic content of the input sentence in a language-independent format. As the analysis in this dissertation suggests (following other studies in cognitive linguistics), language expressions do not *directly* reflect objective events and situations in the world, but rather linguistically express *partial* aspects of the communicated event, which in turn trigger the reconstruction of additional semantic content through cognitive operations such as mapping and blending. Translation (as suggested in chapter 8) *proceeds from these elaborated semantic representations*, rather than from the partial information provided in the input text.

9.1.2.1 Corpus-based (empiricist) methods

The MT methods described so far are all *rule-based* (i.e., they make use of an explicit set of structured symbolic rules to define the linguistic changes involved in the translation of a source sentence into another language). The rule-based systems are contrasted with *empiricist* systems which recently gained new popularity. In empiricist systems, knowledge acquisition and processing is based on *statistical* methods rather than logical rules. The hope is that whatever knowledge is needed for the NLP system will be derived by statistical examination of real texts rather than being coded by human experts and deduced by rules. The assumption is that much linguistic knowledge is acquired and used by statistical and pattern matching techniques on previous observations⁷.

⁷ Note that, from a *cognitive* point of view, while the statistical linguistics assumption (that much of linguistic knowledge is acquired by statistical generalizations over previously heard instances) seems extremely plausible for describing one's own native *language acquisition*, the assumption does not seem as plausible when considering *translation performance*. Speakers are not typically exposed to large streams of discourse that is simultaneously translated (and from which translation examples can be extracted and generalized).

Empiricist MT systems extract the knowledge required for translation from already translated examples. Two basic methods are distinguished (in neither case is there any linguistic analysis in the traditional sense of the source text):

1. *Example-Based* (or Memory-Based) MT - A database of examples (usually aligned bilingual corpora of human translated text) is used to produce new translations by analogy. This method was first suggested by Nagao (1984). In the system presented in Sato and Nagao (1990), the examples are stored as pairs of dependency trees with definition of correspondence links between the source and target nodes. In the process of translation, the input is transformed into a dependency tree and matched with (sub)trees in the database. Using the correspondence links, target dependency trees are created and the target equivalent is eventually generated.

An appealing feature of the example-based MT method, as Kay *et al.* (1994:70) note, is that it can be integrated with a more conventional (rule-based) approach; for instance, the example database might be invoked for difficult constructions (using pre-defined translations); in other cases a conventional transfer or interlingua approach could be used.

2. *Stochastic* MT (statistical model) - In stochastic MT systems, the target sentence is found by a search for a sentence which is the most likely translation of the source (cf. Brown *et al.*, 1990, 1991). This view traces back to methods in information theory, which defines information probabilistically: based on large corpora, probabilities of words are determined given the context (previous words). The basic idea is to regard the occurrence of the target text as conditioned by the occurrence of the source text, and to search for best target text. The algorithm calculates the probability of the source text given various possible target texts, and aims at maximizing this probability function. Probabilities are calculated in advance (from large corpora) for the source language and target language independently (i.e., the probability of a word in a sentence given the previous one or two words), as well as for target words given a source word (calculated from aligned translated texts). The

search algorithm for the translation begins with partial translations of the source text, extending them word by word until there is a complete translation which is “more promising” than any of the other previous candidates.

Corpus-based methods for MT have gained a surprising level of success (relatively to what was expected from such systems⁸), but only few MT researchers today believe that these corpus-based methods will supersede the rule-based methods (cf. Wilks, 1993; Hutchins, 1995). As Kay *et al.* (1992:71) note, corpus-based methods fail in dealing with genuinely ambiguous input (just like traditional methods would with no extra-contextual considerations): in any given corpus containing sufficient occurrences of an examples with several possible translations there will probably be some statistical preference for translating it one way or another. However, the statistically motivated choice will be wrong a large percentage of the time. The reason, as Kay *et al.* emphasize, is that the information telling us how to translate a given input sentence usually lies in the *context* of its use, and it is only by examining features of the context that we can find the right translation. Empiricist MTs may do better by using larger stretches of text (with more context) as the analogizing unit, but then the task of constructing the analogizing database grows unmanageable⁹.

In contrast to the corpus-based approaches discussed above which gained some

⁸ Wilks (1993:5) cites a reported (but unpublished) DARPA-supervised test of the IBM MT system CANDIDE (Brown *et al.*, 1990, 1991) suggesting that the system “did well”, though not as well as the rule-based system SYSTRAN. In another informal report the IBM group claims to get about 40-50% of the translations right (but it is not clear how this number is calculated). It should be noted though that the IBM system is not purely statistical anymore (i.e., it incorporates some “rule-based” linguistic knowledge structures, such as morphology tables, and some use of bilingual dictionary (Wilks, 1993).

⁹ Kay *et al.* (1994) summarize the state of empiricist MT research as follows: “the basic problem with analogizing (empiricists, *N.M.*) approaches is not that they cannot be improved. They clearly can. It is that improving the fidelity of the statistical or example model only promises marginal improvement in the overall performance of a system. There will always be significant problems that fall outside the system’s reach (due to non-local context problems, *N.M.*)” (p.72).

success, no real successful *connectionist* MT research has been reported so far. Connectionism has up to now been used in NLP mainly for parsing and lexical disambiguation. In disambiguation by networks, the activation of a node by an input causes the activation of those other nodes in the network to which the first node is connected. An ambiguous word activates nodes corresponding to all possible senses. If the correct meaning was pre-activated by previously identified concepts, the correct node has a greater activation potential than its 'competitors'. In this way, the wrong interpretation can be eliminated (Cottrell, 1989). The same principle has been used for *syntactic* analysis: dependency constraints may be represented as excitatory or inhibitory links between nodes (Waltz & Pollack 1985). Kay *et al.* (1994, p.79) conclude a short discussion of Connectionism and MT suggesting that "the greatest realm of promise for connectionist processing lies in accounting for preferences", such as preferences between different parsing options of a sentence or readings of an ambiguous word (the choices can be learned in advance with some sort of connectionist weighting scheme). Kay *et al.* also note that this kind of associative weighting can proceed in parallel with conventional linguistic processing.

Some recent papers report on *hybrid* approaches to MT where statistical methods are integrated with traditional AI methods to "fill knowledge gaps until better knowledge bases or linguistic theories arrive" (Knight *et al.*, 1994, p. 134; see also Chang & Su, 1993). Most researchers in the MT field believe that future MT systems will be hybrid, "selecting the best and most effective aspects from both rule-based and corpus-based methods" (Hutchins, 1995:xx; see also Wilks, 1993).

9.1.3 Linguistic problems in the development of MT systems

In this section, I briefly summarize the main types of linguistic problems that developers of MT systems have faced, as reported in the literature.

MT systems, as was noted before, differ mainly in the depth of source text analysis they perform before attempting to generate the target text. An extreme example of MT approach with the least source text analysis is the “Direct Lexical Transfer MT” approach. Lexical Transfer systems attempt to take the most direct route from a sentence in the source language to its equivalent in the target language. That route is determined essentially by two processes: replacement and adjustment. Such a (minimal) system may consist of: (i) a bilingual dictionary that provides potential replacements for each word in the source language; (ii) rules for choosing the correct replacements; and (iii) rules of adjustment for putting words in the right order in the target language, adding or deleting words where necessary, etc.

In order to choose the correct replacement for each word in the source language, Lexical Transfer systems require, however, the extraction of *syntactic* information from the sentence to resolve lexical ambiguities. The following examples display the need for linguistic contextual information (examples from Lehrberger and Bourbeau, 1989):

(i) Homography: when a word belongs to more than one part of speech (this is very common in English as most nouns in English also function as verbs with no change in morphology). In such cases, the lexical form itself may not be enough to define the right translation. Information about the syntactic environment can help in choosing the right part-of-speech.

(ii) Complex constituents when the translation can be obtained only for a whole sequence of lexical items and not by translating each word separately. This is very common, for example, when translating verb and post-particles, as in 2:

(2)	<u>English</u>	<u>French</u>
	pick up	ramasser
	shake up	agiter

Entering the whole sequence (e.g., ‘pick up’) as a single entry to the transfer rules in the system will not solve the problem since the constituents of the sequence may be

separated from each other in the sentence (but still translated as one unit), as in 3b:

- | | | |
|-----|---|-------------------------------|
| (3) | <u>English</u> | <u>French</u> |
| (a) | John picked up the coin. | John ramassa la pièce. |
| (b) | John picked the coin up . | John ramassa la pièce. |

Just searching for the different components (i.e., *pick* and *up*) distributed in the sentence is not enough either, since the components may in fact belong to different sub-clauses, as in 4 below:

- | | | |
|-----|------------------|---|
| (4) | <u>English</u> : | He picked a fight with the guy up the street. |
| | <u>French</u> : | Il en vint aux coups avec le gars de l'autre bout de la rue. |

It is only through *syntactic analysis* of the whole sentence that complex constituents can be identified as a single unit for translation.

Syntactic analysis is required not only to identify the “basic units” for translation, but also to decide on the particular interpretation (and hence translation) of word sequences, as in examples (5-6) below. In 5-6, the semantic properties of the syntactic object define the translation:

- | | | |
|-----|------------------|----------------------------------|
| (5) | <u>English</u> : | She turned on the light. |
| | <u>French</u> : | Elle a allume la lumière. |
| (6) | <u>English</u> : | She turned on the gas. |
| | <u>French</u> : | Elle a ouvert le gaz. |

An additional problem that occurs in lexical transfer systems (with no syntactic analysis) is a problem of duplication of transfer rules: if the contextual constraints are defined in terms of the actual positions of the elements in the sentence, then the system developer must, for example, state twice each translation constraint for a verb and its arguments once for the *active* form and once for the *passive* form.

Probably in all rule-based MT systems today (but not corpus-based systems), the first stage of translation is source text syntactic analysis - the determination (and possible "regularization") of the sentence structure. This stage was also historically the first to be developed by computational linguists.

Syntactic analysis, however, involves additional problems of *disambiguation*. To solve syntactic ambiguities, MT system developers make use of lexical-semantic properties (also referred to as *selectional restrictions*). Consider the following example (from Grishman, 1986):

(6) I noticed a man on the road wearing a hat.

Sentence 6 has two syntactic analyses and correspondingly two semantic representations (one with the man wearing the hat and the other with the road wearing it). If we can determine that 'the road is wearing the hat' is meaningless, we can exclude that reading and home in on the other. This is done by coding (in advance) a “selectional restriction” on the predicate *wear* that its agent must be human (note that selectional restrictions, while helping in restricting the number of syntactic analyses for an input sentence, also prevent the system from accepting *metaphorical* utterances).

For the purpose of MT, many syntactic ambiguities need not be resolved because they can be preserved in the target text as well. Consider the following example 7. The example involves syntactic (and semantic) ambiguity in English (regarding the attachment of the prepositional phrase). The prepositional ambiguity can be preserved in the French translation because both prepositional meanings, the “possession” meaning attached to the 'woman' ('a woman with..') and the “instrumental” meaning attached to the verb ('see with...') can be rendered by the same preposition in the same syntactic location in English and French.

(7) English: I saw a woman on the hill with a telescope.
French: J'ai vu une femme sur la colline avec un telescope.

However, if sentence 7 is translated into Russian, for example, the ambiguity must be resolved since Russian expresses each prepositional meaning differently (Nirenburg *et al.*, 1992:27) .

Few MT systems deal with source text analysis problems beyond syntactic analysis and

the use of selectional restrictions for disambiguation. Nirenburg *et al.* (1992:21-25) discuss additional problems in source text analysis which influence the quality of the translation. These problems go beyond syntactic disambiguation and attachment of lexical-semantic properties to syntactic elements and into the realm of what is traditionally considered “pragmatic” information:

(i) Anaphora resolution: Pronouns like ‘it’ in English are translated differently (in many languages) when they refer to a male or a female. An MT system therefore must assign a particular reference to pronouns in order to choose their correct translation.

(ii) Ellipsis: it is very often the case that elliptical fragments must be translated in full in the target language (and hence must be recovered in the source text).

(iii) Metaphor and metonymy understanding: Nirenburg *et al.* (1992) refer to work by Lakoff and colleagues demonstrating that metaphors are not reserved for poetic texts but are prevalent in everyday language. Nirenburg *et al.* note that an MT system must know (p.25) “whether the systems of metaphorical comparisons among languages are similar and whether they can be translated directly” (see also discussion in section 8.4.6 of this dissertation).

Problems in translation also arise from “mismatches” between the source and target language lexicons (see section 8.1.4). For example, the verb *run* in English can be used with different subjects to express different extensions of the prototype sense of ‘run’. In other languages, each sense will be translated differently (i.e., the equivalent of ‘run’ in the target language may not be extended in a similar way as in English). The following examples are from Larosn (1984:7):

(8)	<u>English</u>	<u>Spanish</u>
	The boy runs	El nino corre (runs)
	The motor runs	El motor funciona (functions)
	The clock runs	El reloj anda (walks)
	His nose runs	Su narriz chorrea (drips)

The problem for the MT system developer is, of course, in defining in advance a list of all possible uses of the word 'run' and its various translations, and identifying the right context for each translation. The experience shows that once a list is constructed, a new use comes up which requires a completely different translation. This is because people creatively extend the meaning of words all the time, and it is of course a deep problem for a computer to recognize such novel extensions, and even worse, to translate them (see further discussion in this chapter).

The deepest problem for MT occurs when the translation must take into account the *general context* beyond the linguistic text - i.e., the situation in which the text was uttered, the intended audience and the culture. Melby (1995) gives as an example the translation of the English expression *thank you* into Japanese. There are several translations and they depend on factors such as whether the person being thanked was obligated to perform the service and how much effort was involved. Even for a human (non-Japanese), it takes substantial effort to learn these distinctions. For a computer, it is impossible to learn. Nagao (1989) similarly discusses the many words of respect and politeness in Japanese which reflect the social position of the speakers, but are hardly used in European languages. Even when these factors are not explicitly expressed in the source European language, they must be inferable from the context and from the psychological state of the speaker, when translated into Japanese.

In the coming sections, I will discuss the problem of translating instances of English Caused-Motion sentences (analyzed in chapter 8) within rule-based MT systems. Note that the translation of the English caused-motion sentences in chapter 8 does not pose any of the classical translation problems discussed above: the sentences are simple to parse and do not involve *syntactic* ambiguity (of course, if the system does not recognize the existence of a Caused-Motion syntactic construction, as suggested by Goldberg, 1995, then parsing difficulties will arise when the system encounter intransitive verbs such as 'sneeze' or

'laugh' occurring with a direct object. However, this problem can be easily fixed by encoding a special rule in the system to deal with this construction. Once such a rule is encoded, the parsing of the examples is quite straightforward.). The translation of the English Caused-motion examples in chapter 8 also does not involve *lexical* ambiguity of individual lexical items (for example, in the sentence *Frank sneezed the napkin off the table*, the information associated with each lexical item in the sentence, and its translation, is the *default* one. That is 'sneeze' refers to the default act of 'sneezing', 'napkin' refers to a prototypical napkin, and so on...). Finally, the translation of English caused-motion sentences into Hebrew or French is not a function of *cultural* differences: most of the translation examples discussed in chapter 8 communicate everyday events which are culture-independent.

The problem of translating the English caused-motion sentences analyzed in chapter 8 is rather the outcome of the *creative* linguistic *combination* (or *blending*) of conventional lexical items and syntactic forms. The problem posed to the (computational) translator results first of all from the need to reconstruct the novel (creative) linguistic blend performed by the speaker. Furthermore, the translator needs to infer additional knowledge necessary for translation but not provided in the source text (or in its larger textual context for that matter). And finally, the translator must blend again the constructed complex semantic structure into a basic clause construction in the target language (rather than directly transferring linguistic units from the source text into corresponding expressions in the target language). In the next section, I will discuss the extent to which current methodologies of MT can deal with these type of cognitive-linguistic operations.

9.2 Implications of grammatical blending for semantic analysis in rule-based MT

In this section, I proceed to the core discussion of the chapter: the use and manipulation

of encoded world knowledge structures for semantic analysis and translation in MT systems. I will suggest that current prevalent methods in NLP for manipulating world-knowledge and semantic structures are not equipped to deal with linguistic blends of the type presented in this dissertation. I will start by discussing approaches to semantic representation in rule-based MT systems.

9.2.1 Meaning representation in NLP

Katz & Fodor (1963) initiated a tradition in Linguistics, Philosophy, and later in AI of semantics as manipulation of *semantic markers* attached to lexical items. This semantic theory posited binary markers (such as, +/- *human, male, animal*) that would be used to build the possible senses of every word. To decide the meaning of any word in a given sentence, a postulated body of rules would describe how these markers could permissibly interact in non-anomalous sentences. Semantic markers combined with syntactic markers and rules of combination would provide us with the meaning of a sentence. Though this approach has been frequently attacked, it is still very influential in the AI/NLP community; Being able to implement semantics using a limited number of markers is computationally very attractive.

The initial idea of binary markers was expanded to *slots and frames* (Minsky, 1975) in the Artificial Intelligence (AI) community. Instead of simply possessing a marker, each lexical entity could contain slots in which were found either a value, a pointer to a default value, or a procedure that supplies the missing value. Hayes (1985) suggested that frame representations could be seen as a new syntax for first-order logic: the frame is a bundle of properties which are instantiated in particular individuals and situations. Each frame instance denotes the individual and each slot denotes a relation which may hold between that individual and some other individuals. Rather than storing assertions in a clausal form, they can now be stored in frames.

Schank (1975, 1977) combined the slots and frames idea with the linguistic tradition of case grammars (Fillmore, 1968) in what he calls *Conceptual Dependencies* (CD). In the CD semantic representation, verbs are described by semantic primitives (e.g., *transfer (physical or mental), move, speak, build*). Case relations with different nouns (that could be agents, patients, instruments, locatives, and so on) are marked for the verbal primitives. The CD representations were used to choose word senses, and represent *scripts* or stereotypical sequences of actions (Schank & Abelson, 1977)¹⁰. The text analysis was done mainly by filling up CD structures.

In Schank's theory a predetermined set of possible relations (conceptual rules) is used to *predict* conceptual items implicit in a sentence. The language analyzer ('conceptual analyzer') makes use of skeletal semantic structures to guide the analysis. These skeletons specify the primitive actions and the type of objects involved, and have places for filling in specific instances of objects involved in a particular event. NLP systems developed by Schank were based on *semantic-driven analysis*. In opposition to the *linguistic* approach, where there are two levels of representation - syntactic and semantic - and constraints on acceptable structures at each level, Schank suggested that since our ultimate objective is the generation of a semantic representation of the sentence, we should do so directly, and use semantic constraints to guide the process. Schank's system analyzes the text directly into semantic structures, called *Conceptual Dependency* networks.

The strongest argument of those advocating semantics-driven syntax analysis has been the ability of people to interpret sentences from semantic clues in the face of syntactic errors or missing information (as in *he go movies yesterday*). The early analyzers developed by

¹⁰ The *scripts* are frame-like structures for representing typical or expected sequences of events. A well known example of a script is the RESTAURANT script, which details the sequence of events and expected behaviors when going to a restaurant (i.e., entering, being seated at a table, being shown a menu, ordering from a waiter, etc.).

Schank and his students begin by identifying the 'main noun' and 'main verb' of the sentence and building an incomplete semantic structure. The sentence is then searched for words of the proper class to complete this structure. This approach should therefore be able to handle ungrammatical sentences which would cause a syntactic analyzer to fail. The semantic-driven analysis, however, also had many difficulties: First, the merging of syntax and semantics made it difficult to capture syntactic generalizations, such as the relation of active and passive forms. Second, analyzing complex sentences, particularly those involving conjunctions and comparatives, with a semantic analyzer is very complicated without some syntactic guidance. Finally, the identification of semantic primitives for semantics-driven systems is extremely problematic and it is not known whether such systems can remain stable with large vocabularies containing several thousand lexical entries¹¹ .

Today, almost all MT systems start their processing of input sentences with *syntactic* analysis¹². *Semantic* properties and frame-like “world-knowledge” structures are typically used to guide the syntactic parsing process mainly, and (in some systems) also to build some level of semantic representation for input texts. The central debate today is between those linking semantic information *directly to lexical items*, and those who advocate a distinction between what is considered to be *linguistic* knowledge (associated with the lexical items in each language), and *extra-linguistic* (“*common-sense*”) *world knowledge*

¹¹ Many linguists and philosophers have argued that the existence of a set of truly universal primitives is unlikely. The problem is deciding which concepts are the “primitive” ones, if any exists at all (for a discussion of some of the arguments for and against semantic primitives, see Y. Wilks, 1987).

¹² Many models of syntactic formalisms have been explored in NLP research, including Transformational Generative Grammar, Categorical Grammar, Lexical-Functional Grammar, and Head-Driven Phrase Structure Grammar.

which is defined independently of any language¹³. The latter type of knowledge (common-sense language-independent knowledge) is often referred to in the MT literature as *ontological* knowledge (cf. Carbonell, 1978; Wilks, 1979; Nirenburg *et al.*, 1992, 1995; Dorr, 1993).

Schank's basic idea of frame-based semantics plays an important role in the computational representation of ontological knowledge. The ontology first provides "a uniform definition of basic semantic categories, such as objects, event-types, relations, properties, and episodes that become the building blocks for descriptions of particular domains" (Nirenburg *et al.*, 1992:69). These categories are used to define "what 'concepts' exist in the world and how they relate to each other" (Mahesh, 1996:5)¹⁴. In some systems, event-type definitions are also used "to encode past experiences, both actually perceived and reported" (Nirenburg *et al.*, 1992:71), in the form of *episodes* (units of knowledge that encapsulate particular 'remembered' instances of events and objects)¹⁵. In a ("knowledge-based") MT system, the ontological knowledge is used to guide all levels of "linguistic" processing: lexical, syntactic, semantic, and pragmatic processing of both source text

¹³ Nirenburg *et al.* (1992), the developers of the Mikrokosmos knowledge-based MT project, note however that while proposals about the content of *lexical* semantic properties generally avoid the concept of language-*neutral* knowledge, they nevertheless "introduce elements of metalinguistic apparatus [in their proposals] which play the same role as ontology [i.e., general world knowledge]" (p.7).

¹⁴ Mahesh (1996) describes in some detail the construction of an ontology for the Mikrokosmos knowledge-based MT system developed at the NMSU Computer Research Laboratory. The ontology of Mikrokosmos is a directed graph where the nodes are the "concepts" (here marked in capital letters). The "concept" is the 'primitive' computational symbol with well defined attributes and relationships with other concepts. Links between nodes are represented as slots and fillers. Slot names themselves are concepts of the class PROPERTY. PROPERTYs are of two types: RELATIONs and ATTRIBUTEs. RELATIONs map an OBJECT or EVENT to another OBJECT or EVENT while ATTRIBUTEs map an OBJECT or an EVENT to a scalar or literal symbol. The 'slot' is the fundamental "meta-ontological" predicate. Each slot has several "facets": range of values (fillers), default value, salience in the entire concept, and more.

¹⁵ The ontology and episodes are sometimes discussed as two different types of memory: semantic and episodic. For convenience, I will refer to the encoding of both types of knowledge by the term "background world-knowledge" or "ontology".

analysis and target text generation (Nirenburg *et al.*, 1992)¹⁶.

Representing and manipulating ontologies is one of the outstanding research questions of the entire discipline of AI. Leaving aside the problem of scale and cost, the concept of ontology in AI has been criticized on the ground of irreproducibility (based on the claim that no two people would be able to agree on what any particular node or path in the ontology hierarchical structure should look like), and on the basis of cultural-dependency (i.e., that the way an ontology is built necessarily reflects the world-view behind one dominant language). As Nirenburg (1992:70) notes, “Even today, this area of scientific research [constructing ontologies] remains to a large degree, as it has been over 2,500 years, within the purview of philosophy”¹⁷.

My interest in this chapter, however, is not in the *feasibility* of constructing computational ontologies. Rather, I assume for the moment that a complete ontology is available in some form to the language-processing (computational) agent. The question I will ponder in the next section is the following: given a computational system which

¹⁶ Mahesh (1996:41) mentions however that the Mikrokosmos ontology concepts are used to “represent linguistic meaning rather than to make elaborate inferences or carry out non linguistic action”. Based on this distinction, the Mikrokosmos ontology does *not* include “prototypical episodic and procedural knowledge”.

The translation examples discussed in this thesis suggest that episodic and procedural information *is* an integral part of translation. The analysis in this thesis and others (e.g., Nunberg, 1995) suggest that no clear border line can be put between “lexical semantic” knowledge and general (encyclopedic) world knowledge. I will return to this point later in the discussion.

¹⁷ An important question regarding the representation of world knowledge is what type of information should be stored. For example, findings in studies on spatial prepositions and their representation (e.g. Herskovits, 1982; Vandeloise, 1991) point to the fact that world knowledge-bases should not necessarily reflect a “valid” or “logical” view of the real world. Rather they should represent the prototypical conceptualization of the world, as reflected in the way people speak. Herskovits (1982) who investigated English spatial expressions found out that “our words often describe mental maps, which are made of lines and points approximating the canonical view of the world” (p.64). According to Lakoff (1982) the central aspect of our language is experiential: mental imagery, memory, and gestalt perception, all have to do with “human interaction with and functioning in the world, rather than with objective properties of the world” (p. 22). It is probably this experiential information that has to be coded as “common-sense” knowledge (see also Mandelblit, 1992, on the translation of spatial prepositions).

Note however that for the purposes of constructing an NLP system, the *psychological reality* of the ontology, or its correspondence to mental structures in the human mind is not of interest (for NLP purposes, the ontology is just a working tool to provide better performance).

represents knowledge about the world in some form, how would the system *manipulate* this information to construct the representational structures required for language processing (and, in particular, translation)?

9.2.2 Manipulation of general knowledge structures for constructing semantic representation of texts

Nirenburg *et al.* (1992:73) describe the role of the ontology in the process of constructing a semantic representation for input texts¹⁸ :

¹⁸ This quotation refers only to the representation of “*textual*” (T) meaning (or semantics). Nirenburg *et al.* also discuss in their manuscript the representation of the speaker’s *goals* (G) and the *setting* (S) of the communication situation as part of the text’s meaning. The text meaning is “a triad $SM = \{T, G, S\}$ ” (1992:75).

During analysis, ontological structures are instantiated in working memory that capture the actual knowledge necessary to "understand" a text or to produce a text or a turn in a dialogue. We believe there is a well-defined set of knowledge elements whose existence constitutes a necessary and sufficient condition for a text to be considered understood... Basically, we represent the semantic content of natural language utterances by instantiating ontological entities or reasserting remembered instances of such entities that are found...to be the most closely semantically related to lexical units in the input.

How are linguistic and ontological structures manipulated to construct semantic representation of texts? Looking at the general NLP literature, we find that rule-based NLP (and AI in general) basically studies only one type of knowledge manipulation: *reasoning* from general knowledge to specific cases, or what is generally referred to as *common sense reasoning*. The discussion is confined to restricted circumstances where *logical inferencing* can take place; i.e. reasoning that involves premises and conclusions, and is based on laws of logic. Understanding a sentence, in this view, is finding a way to make it true. Semantic representations are assigned to the parts of a given sentence, so that given the ontology (or *model*), one can tell whether the sentence is true in that model. The paradigm cases used for inferencing are *deduction*, *induction*, and *abduction*¹⁹.

Note that there is an important implicit assumption underlying the logical reasoning manipulation of knowledge employed in NLP systems: the assumption that there is a set of rules and world-knowledge frames that if comprehensive enough can predict and give a model to all future language generation and understanding via processes of logical inferencing. In other words, the assumption is that the 'mental' structures that represent the interpretation of any text *either exist as such in the system in advance, or can be derived by*

¹⁹ Kay *et al.* (1994) describe these three forms of inferencing: In deduction, from $(\forall x)p(x) \rightarrow q(x)$ and $p(A)$, one concludes $q(A)$. In induction, from $p(A)$ and $q(A)$, or more likely, from a number of instances of $p(A)$ and $q(A)$, one concludes $(\forall x)p(x) \rightarrow q(x)$. Abduction is the third possibility. From $(\forall x)p(x) \rightarrow q(x)$ and $q(A)$, one concludes $p(A)$. That is, $q(A)$ is seen as observable evidence, where $(\forall x)p(x) \rightarrow q(x)$ is a general principle that explains $q(A)$'s occurrence, and $p(A)$ is the inferred underlying cause or explanation of $q(A)$.

logical inference rules from other structures in the system.

The point I would like to make in this section is that the basic mechanisms of semantic processing discussed above - i.e., the construction of semantic representation of input texts by *directly instantiating* the semantics of the linguistic structures in ontological frames (or their logical derivations) cannot support *all* forms of language processing. In particular, the semantic content of linguistic utterances such as the "blends" discussed in this dissertation is not prototypically a direct instantiation of any individual ontological structure (or its logical derivation), but rather an "instantiation" of *partial* information from *several* independent ontological structures linked (or integrated) together by unpredictable analogical mappings. In the next section, I will discuss as an example the construction of semantic representation for English Caused-Motion sentences.

9.2.3 Constructing semantic representation from ontological frames for English Caused-Motion Sentences.

Consider the following input text to an NLP system:

(9) The audience laughed the actor off the stage.

In chapter 2, the linguistic and conceptual blending processes underlying in the generation of sentence 9 were discussed (Following Fauconnier & Turner, 1996): various linguistic structures (lexical forms and grammatical constructions which conventionally represent events and relations in the world) are integrated and form a *temporary* structure - the 'blend'²⁰. The 'blend' is reflected in the actual utterance in the language and is the *input* for an NLP system to analyze. Given an input text such as 9, from the point of view of NLP system developers, the question is: which ontological knowledge frames need to be encoded in the system in advance and how these structures are to be retrieved to provide a

²⁰ By 'temporary', I refer to the fact that the linguistic blend is often created for the purpose of only one text or even one sentence.

correct semantic model for sentence 9?

Note that in NLP systems the semantic representation of clauses is typically constructed as a representation of *predicates and their arguments* (corresponding to the sentence's main verb and nominal phrases). Semantic analyzers typically begin by identifying the main verb of the sentence and retrieving an ontological frame that represents the “semantic predicate” associated with the verb (the ontological frames are linked to verbs in the lexicon). In the ontological frame, “case relations” (agents, patients, instruments, locatives, and so on) are marked for the predicate. The semantic processing of the sentence involves the attachment (or “linking”) of nominal and adverbial components in the sentence to the “case relations” slots in the ontological frame. In discussing possible ontological frames to represent the semantics of input sentences in this section, I will therefore focus on ontological frames to be retrieved by the sentences' main verbs (e.g., the verb *laugh* in example 9).

One option for constructing a semantic representation for input sentence 9 based on ontological frames is to have in the ontology a single frame which represents a single conceptual predicate meaning “cause to move by laughter” (I would term this frame LAUGH-CM). Note that this is an acceptable option if we base the construction of the ontology and its links to the on-line English lexicon on conventional (paper) English dictionaries. The Webster ninth Collegiate Dictionary, for example, defines one of the senses of the English verb ‘laugh’ (in its transitive use) as “to influence or move by a laughter”. The frame LAUGH-CM would then include several slots for the various nominal participants (“case relations”) associated with the predicate: an agent, a moving patient, and a source/goal location. Linking rules would attach the linguistic nominal arguments of the verb *laugh* in the input sentence 9 to participant slots in the ontological frame (i.e., the subject would be linked to the agent slot, the object to the patient or theme slot, and so on...). The frame representation for LAUGH-CM would look something like Figure 9-1

below ²¹. Each filler may itself be an ontological frame or a ‘literal’ value (numerical or alphabetical). In brackets are the "fillers" for the frame LAUGH-CM for input sentence 9.

LAUGH-CM	Def: "A move P (from S to G) by laughter "
Agent (A):	(audience)
Patient (P):	(actor)
Source (S):	(stage)
Goal (G):

Figure 9-1: A frame-type representation for the predicate LAUGH-CM.

Note that the frame in Figure 9-1 represents, in fact, a *conceptual integration* of the sort argued for in chapter 2. That is, a *sequence* of events in the real world (i.e., Agent laughs, Patient moves) is defined as a *single* integrated ontological concept. The *structure* of the frame (i.e., the case relations associated with the predicate, and the mapping of grammatical roles to semantic case relations) is the same as the ones for caused-motion concepts such as THROW or PUSH (the concepts associated with the English lexical items *throw* and *push*).

The problem with this approach is, of course, that it cannot account for *creative non-entrenched* instances of the English Caused-Motion construction (i.e., creative linguistic *integration* of conceived caused-motion event sequences into a single linguistic construction), as in 10:

(10) Frank sneezed the napkin off the table.

Clearly, we cannot expect an ontology in an NLP system (however rich and detailed it is) to include a single frame representation for the *integrated* caused-motion semantics of sentence 10 (i.e., a frame which represents a predicate with the semantics of “to move by *sneezing*”). If we do, then we must define in the ontology a *second* frame for every non-stative verb *V* in English, with the semantics of “to influence or move by *V*” (since each

²¹ The representation format in Figure 9-1 is a simplification of figures such as the ones found in Nirenburg *et al.* (1992) or Mahesh (1996). Their format is based on the FRAMEKIT knowledge representation system (Carbonell and Joseph, 1985).

non-stative verb in English can *potentially* be integrated into a CM syntactic structure to represent a caused-motion event sequence, as the study by Goldberg, 1995, suggests).

Another option would be to construct a representation of the semantics of sentences 9 or 10 as instantiations of a more generic frame in the ontology, representing the generic event structure of CAUSED-MOTION (i.e., a frame representing a generic recurring event sequence in the world of 'an Agent acting and thereby causing a Patient to move'). The event that the frame represents would be associated with the same "case relations" as the frame LAUGH-CM (Figure 9-1), but it would include an additional slot identifying the particular type of activity involved in each instantiation of the frame (e.g., 'laughing' vs. 'sneezing' in 9-10). Figure 9-2 provides a schematic illustration of the frame CAUSED-MOTION²²:

CAUSED-MOTION	Def: "A causes P to move by means of act Ac"
Act (Ac):	(<i>laugh / sneeze</i>)
Agent (A):	(<i>audience / he</i>)
Patient (P):	(<i>actor / napkin</i>)
Source (S):	(<i>stage / table</i>)
Goal (G):

Figure 9-2: A frame-type representation of CAUSED-MOTION events.

Note that the grammatical linguistic encoding of a caused-motion event in English (in a single clause structure with a single verb) cognitively motivates the representation of this category of events as an independent ontological frame ("concept"), rather than, say, as a complex combination of several frames representing the different sub-events in the macro caused-motion event). Note also that from a pragmatic computational point of view, the

²² The semantic representation of the concept "caused-motion" in Figure 9-2) is clearly partial. For example, as the Caused-Motion syntactic construction in English reveals, the *direction* of motion (up, down, into...) is a salient aspect of the caused-motion event (in language and probably in conceptual perception as well), and should therefore be an integral part of the semantic/conceptual representation of caused-motion events in the ontology.

latter representational option (Figure 9-2) is more efficient than the earlier one (Figure 9-1). That is, rather than constructing (in the ontology) two frames for each non-stative verb in English such as *laugh* or *sneeze* -- one for the causative sense and one for the non-causative 'basic' sense of the verb -- the causative sense can be *derived* from the frame representing the non-causative sense plus the generic frame of CAUSED-MOTION. In analyzing English sentences, the CAUSED-MOTION frame (Figure 9-2) will be triggered by the *syntactic* pattern of the English input sentence [NP V NP directional-PP], and the main verb will identify the particular 'type' of activity involved in the generic event structure of caused-motion.

The problem, however, is that for the purpose of translation, for example, the semantic representation of English Caused-Motion sentences as an instantiation of the ontological frame in Figure 9-2 is insufficient (i.e., is not "functionally complete"). As the discussion of English CM sentences (in chapter 2) and their translation into Hebrew and French (in chapter 8) suggests, the main verb in the English Caused-Motion construction (represented by the generic 'activity' slot in the frame in Figure 9-2) may refer to *different sub-events* within the conceived caused-motion macro-event, and which sub-event the verb refers to defines the *translation strategy* into the target language (i.e., the choice of the integrating syntactic construction and lexical items in the target language, see discussion in chapter 8). Moreover, as the analysis in chapter 8 suggests, what the main verb in the English caused-motion sentence refers to also defines what essential information is *missing* from the source sentence linguistic structure, information that is essential for successful translation. For example, if the verb in the caused-motion sentence depicts the *causing* sub-event, as is often the case in English, then for translation purposes into Hebrew and French, as well as into many other languages, the *effected* motion event must be *inferred*, since this latter aspect of the event is the one most commonly highlighted by the constructions of the target languages (see section 8.4.1).

Therefore, for the purpose of translation, the “interlingual” semantic representation of English Caused-Motion sentences must first be *extended* to distinguish between (at least) three possible predications the main verb in the English caused-motion sentence may denote (i.e., the agent’s action, the patient’s motion, or the causal link between the two²³). Second, for correct translation of CM sentences into Hebrew and French, the particular *type* of effected motion event must typically be represented as well. Figure 9-3 depicts what would be a “sufficiently complete” representation of the semantics of sentence 10 for translation purposes into Hebrew and French (below, is the translation of sentence 10 into Hebrew, as discussed in section 8.4.1):

- (10) English: Frank **sneezed** the napkin off the table.
Hebrew: *Frank hepil(n.f.l-hif'il) et hamapit min hashulchan behitatsho.*
Frank **fall-hif'il**_{past} ACC the-napkin off the-table **by-sneezing**.

CAUSED-MOTION(extended) Def: “AG causes(C) PA to move(M) by means of (AC)”	
Predicate:	
Causing Activity (AC):	<i>(sneeze)</i>
Effected Motion (M)	<i>(fall)</i>
Causal Link (C):	
Agent (AG):	<i>(Frank)</i>
Patient (PA):	<i>(napkin)</i>
Source (S):	<i>(table)</i>

Figure 9-3: A (partial) frame semantic representation of the sentence *Frank sneezed the napkin off the table*

The problem for computational processing is that the construction of the frame representation as in Figure 9-3 *cannot be derived from other ontological frames* in the system *by logical inferencing only* (i.e., by operations of induction, deduction, or

²³ The main verb in the Caused-motion construction may also denote other aspects of the event, as exemplified in Goldberg, 1995, Fauconnier & Tuner, 1996. and in section 8.3.4.5 in this dissertation.

abduction).

First, the frame representation in Figure 9-3 cannot be derived from the generic CAUSED-MOTION frame (Figure 9-2) because the grammar in English underspecifies the correct mapping rule between the sentence's main verb and the three (minimal) optional "predication" slots in Figure 9-3²⁴. Second, the missing information (the manner of motion of the affected patient) cannot be derived from any single ontological frame (associated with individual lexical items in the sentence) such as SNEEZE, NAPKIN, or TABLE. Nothing in the properties of the concept 'sneeze' or 'napkin' alone suggests that in a caused-motion sequence, the effect of sneezing on a patient would necessarily be one of 'falling' (as in the representation in Figure 9-3), rather than, say, 'shifting aside', or 'running away'. Consider for example CM sentence 11, with the same main verb *sneeze*:

(11) Eng. Frank **sneezed** everyone out of the room.

The effected motion event associated with sentence 11 is probably not one of 'falling' but rather of volitional 'walking' or 'running out', as reflected in the translation below of sentence 11 into Hebrew (provided by two Hebrew speakers):

(11a) Heb. *Frank hivriax(b.r.x-hif'il) et kulam min haxeder bahit'atshuyot shelo.*

Frank **run away-hif'il**_{Past} ACC everyone from the-room **by-his-sneezing**To infer the effected type of motion event, in 10 or 11 the language user must infer a probable causal link between the agent's action ('sneezing') and the motion of the patient.

²⁴ Note that the semantic properties of the main verb in English caused-motion sentences can often *cue* the system as to which aspect of the event the verb depicts (i.e., verbs of motion most probably depict the patient's effected motion; otherwise the verb most probably depicts the agent's causing activity, and so on). However, these semantic properties are not always sufficient, since the same verb may be used to indicate either the agent's action, or the object's motion, as exemplified in sentences (i-ii) from Fauconnier and Turner (1996):

- (i) *He trotted the stroller around the park*
- (ii) *The trainer trotted the horse into the stable*

In (i), the verb 'trot' refers to the causal agent (it is the causal agent who is doing the trotting, and thereby making the stroller move around the park). In (ii), the verb 'trot' refers to the affected patient (it is the horse trotting into the stable. The trainer could be walking, holding the horse's bridle).

While in 10, it is probably the air displaced by sneezing that *physically* caused the napkin to move, in 11 the effected motion is assumed to be volitional and is due to cognitive (epistemological) decision of the patient rather than the physical causal effect of sneezing. The different properties of the affected patient and the difference in the implied causal force suggest different manners of effected motion events in sentences 10 and 11. In other words, to infer the manner of the effected motion event, *as required for translation*, a whole sequence of events must be reconstructed, partial aspects of which link to lexical items in the linguistic utterance. The MT system, thus, must perform the full "de-integration" and blending reconstruction operations discussed in the thesis (section 2.4) for generating a semantic representation of the input sentence which is "sufficiently complete" for translation. The association of a linguistic utterance (such as 10-11) with a sufficiently complete semantic representation is therefore clearly not an operation of logical *instantiation*. Rather, the input sentence (the linguistic blend) must be linked to *fragments* of episodic memory (e.g., memories on the typical motion of light objects like 'napkins', of observed physical effects of 'sneezing' on objects around the sneezer, etc.). Each episodic fragment is encoded separately in the ontology and all are combined together only *temporarily* for the specific linguistic blend (input sentence).

Note that the above discussion is also relevant for more frequently observed causal scenarios (whose integration into a single structure is more entrenched in language and memory) as in example 9 discussed before:

(9) The audience laughed the actor off the stage.

In the ideal case (i.e., given an ontology as rich as the human memory), what the ontology might include is a frame in episodic memory representing a full sequence of events whereby a group of people laugh at another person (or more generally perform some insulting act towards another person), and the laughter (insult) makes the person leave the place where the group is located. This general episode is probably common enough to be

encoded in the human mind (i.e., most people have probably encountered such as episode either by being themselves a ‘victim’ of social insult, or by watching, reading or discussing such an event and its causal effects). This event may be represented in the ontology as follows (Figure 9-4):

CAUSED-MOTION-by-insulting Def: “A insults P, thereby causes P to move away”	
<i>Event-1:</i>	
Insulting-act (I):	(<i>laugh</i>)
Agent (A-I):	(<i>audience</i>)
Patient (P-I):	(<i>actor</i>)
at-Location:
<i>Event-2:</i>	
Motion-act (M):	(<i>laugh</i>)
Patient (P-M):	(<i>actor</i>)
from-Location:
to-Location:

Figure 9-4: A frame representation of the generic event structure
MOTION-CAUSED-BY-INSULT.

Note however that the episode (or category of episodes) represented in Figure 9-4 may also be encoded linguistically in many ways, as in “X laughed Y out of Z” or “X made Y leave Z (by insulting her)”, or “Y ran-away from Z (because of X’s insult)” and so on... Each linguistic expression of the episode involves a different *linguistic blending* of the sequence of events into a different syntactic construction while *highlighting* and *omitting* different aspects of the event (i.e., each integration involves a different *mapping configuration*). In other words, even when a whole recurring communicated sequence of events is already represented as a single frame in the ontology, there is no one well-defined "instantiation rule" between the frame structure and the linguistic form. The language interpreter (the NLP system) must still infer the particular mapping configuration involved in each linguistic blend.

To summarize, the discussion in this section points out that the computational translation of some basic clause structures (even isolated sentences with no background cultural or textual consideration) may require (for the very basic “functionally sufficient” semantic representation of the sentence) not only a detailed episodic memory, but also sophisticated computational tools to discover the *links* between the linguistic utterance and episodic frames in memory. The discovery of these links involves computational mechanisms which often go beyond simple logical inferencing from general frames to instantiating events, involving rather the reconstruction of complex mapping patterns from fragments of episodic memory to the linguistic form, where some aspects of the event are linguistically represented and others are not explicitly expressed. To provide the right interpretation for each possible linguistic blend using only methods of logical *instantiation* (or inference from general case to particular instances) would involve the *encoding* of *each possible linguistic integration* (blending) from each set of event structure to several appropriate linguistic structures. This means in fact that all possible blends that can be generated by human speakers in any time must be encoded (either each one individually, or for more general categories of very similar events, i.e. events which involve very similar agents, patients, and interactions). This, of course, is not feasible.

It is important to note here, however, that the discussion in this chapter refers to translation examples where the linguistic blend underlying the expression of an event in the source and target languages is *different* (as dictated by the constraints of the different grammatical constructions in the two language). While the tenet of this dissertation, as discussed in chapter 2, is that every linguistic utterance is generated by linguistic blend (i.e., by the mapping of an event onto an integrating syntactic construction), it is not always the case that a full reconstruction of the blending configuration is needed for the

purpose of translation, since in many cases the *same* linguistic blend is conventionally employed in the source and target languages. In such cases, translation can proceed by simply replacing lexical items (or sub-clauses) in the source language with their dictionary equivalents in the target language. The source and target language readers will each independently reconstruct the full communicated event from the partial information in the linguistic structure (the source or target sentence).

I suggest that similarities in blending configurations and conventions in the source and target languages is what enables successful automatic translation (by MT systems) of many linguistic utterances, and it is also probably what underlies the simplistic engineering view of translation (discussed in section 9.1.1) as a direct “decoding” process. It is however important to note that while many sentences can be translated into a target language *without* figuring out the complicated mapping between the linguistic structure and episodic/ontological frames (only because the same mapping configuration is used in both the source and target language), in other cases reconstruction of the blending operation is required.

From a practical point of view, the importance of the discussion in this chapter for current NLP systems is in the way NLP developers treat cases where NLP systems *fail* to produce the right output (e.g., cases where direct transfer of linguistic units into the target language does not produce an acceptable translation). The almost automatic assumption in such cases is that the failure stems from *incomplete knowledge* encoded in the system (i.e., the linguistic knowledge and algorithms, the ontological frames, or the statistical patterns retrieved from corpora are assumed to be imprecise or not complete). The reaction of NLP system developers to failures of the system is thus usually in a *modification* of the knowledge structures in the system (too often in ad-hoc fashion) just to discover that another problem arises in the next round. What the discussion in this chapter suggests is that the knowledge-bases themselves may be correct and complete, while it is the *creative*

way these permanent knowledge-structures are *integrated* (conceptually and linguistically) that is not captured by the system's algorithms²⁵.

In the next section, I will briefly discuss some current research trends in NLP to extend the power of NLP systems (preventing failures in processing input texts). These methods focus on the role of *lexical* properties in defining the semantics (and translation) of a sentence, and try to cope with failures of the system by enhancing the lexicon with general *sense-extension rules*. These methods, however, fail again to acknowledge and treat the real creative aspect of language processing. While the general rules encoded in these methods can capture *conventional* blends (i.e., blends that are entrenched and repeated in conversation), they are not general enough to account for temporary novel blends.

9.2.4 Lexical mechanisms in semantic processing.

A large number of recent NLP systems encode semantic information as “lexical-semantic properties” associated directly with lexical items (rather than in the form of ‘metalinguistic’ ontological knowledge). This follows a general current move in linguistics theories to attempt to account for all semantic interpretation and surface syntactic structure through lexical-semantic information encoded in individual lexical items (e.g., Bresnan, 1982; Levin, 1993).

Recent work has increasingly emphasized the “creative” aspects of language use where *word* senses can be *extended* in context in cases such as metonymy (Nunberg, 1978, 1993), and metaphor (Lakoff and Johnson, 1980). These phenomena are often treated in

²⁵ For example, in the sentence *Frank sneezed the napkin off the table*, the information associated with each lexical item in the sentence is probably a default one (that is ‘sneeze’ refers to the default act of ‘sneezing’, ‘napkin’ refers to a prototype napkin, and so on). The novelty of this expression is in the way these default knowledge structures are linguistically integrated together, and hence related to each other semantically in the sentence. Modifying the default properties of ‘sneeze’ or ‘napkin’ to provide the correct interpretation would be only an ad-hoc solution and will not solve the general problem.

computational linguistics literature through lexical mechanisms such as *coercion* (Pustejovsky, 1991; Pustejovsky and Boguraev, 1993), and *lexical extension rules* (Copestake & Briscoe, 1995). In particular, research has concentrated in recent years on identifying classes of lexical items (with special reference to verbs) with shared semantic components, whose syntactic “behavior” can be predicted from the meaning components.

For example, Levin (1993) discusses verbal diathesis alternations - alternations in the expressions of arguments, accompanied by changes of meaning. The diathesis alternations are assumed to arise systematically from the verb meaning. Therefore, groups of verbs which form a semantically coherent class would share the same alternation patterns. Levin notes, for example, that most verbs of sound emission (e.g., *buzz*, *hiss*, *rattle*, *wheeze*, *whistle*) allow locative (or directional) alternation, as in 12:

- (12) *The wind whistled.*
The wind whistled round them.

This observed behavior supports a general “sense extension” rule as in 13, which is accompanied by an alternation in argument structure (i.e., the addition of a locative prepositional phrase).

- (13) Extension Rule: sound emission V --> motion V

In Copestake (1995) and Ostler *et al.* (1992), sense extension rules for nouns are discussed, such as:

- (14) (a) ‘Animal (countable noun) -> Meat (mass noun)’
*Marry had a little **lamb**.*
*He won’t touch **lamb** anymore.*
- (b) ‘Comestible substance -> Conventional portion’
*She doesn’t drink **beer**.*
*She bought two **beers**.*
- (c) ‘Container -> Its content’
*The **bottle** broke.*
*He drank a **bottle** of Whisky.*

Work on other languages suggests that sense extension rules (such as 13-14) are found

cross linguistically, but may differ in the groups of lexical items to which the rules are applied, and the syntactic and morphological phenomena involved in the alternation. It follows then that the application of sense extension rules across languages may lead to translation divergences (or mismatches).

For example, while *walk* is translated into French as *aller* (or *aller à pied*), the phrase *walk across* is translated as *traverser à pied* (i.e. ‘cross by foot’). This “translation divergence” could be accounted for by *extending* the sense of *walk* in English into a sense of ‘crossing’ (in the context of the adverb *across*), an extension rule which is not paralleled in French. In the same spirit, following Talmy’s work on motion event conflation (Talmy, 1985, 1991), the *motion* semantic component in sentences like *the bottle floated into the cave* could be incorporated by extending the semantic properties of the verb *float* to include a meaning component of ‘directed motion’. The sense-extension rule will solve the problem of translating this sentence into “verb-framed” languages (in Talmy’s terms) which normally express the direction of motion (the ‘path’) in the verb, while information about manner (‘floating’) is expressed separately (or just omitted). Thus, for example, the fact that the English sentence *the bottle floated into the cave* is translated into Spanish (a “verb-framed” language) as *la botella entró a la cueva* (‘the bottle *entered* into the cave’) can be captured by a sense-extension rule for the verb *float* (which adds the meaning component ‘directed motion’, and a transfer rule which links the extended sense of *float* and the verb *entrare* in Spanish through the shared meaning component of ‘directed motion’). Another related solution is proposed by Dorr (1993), in a comprehensive study of a large bulk of translation divergences. Instead of solving translation divergence problems through sense-extension rules, Dorr proposes to treat them by identifying general surface-level distinctions across languages at the level of lexical-semantic structure, and factoring out these differences in the translation process.

Note that following this line of research, the “translation divergences” discussed in

chapter 8 (translating English Caused-Motion sentences into Hebrew and French) would also be treated by pre-encoded rules at the *lexical* level. For example, the “translation divergence” in 15 (below) may be accounted for by encoding a rule which *extends the sense* (or lexical properties) of the English verb *blow* (or its larger lexical-semantic category) to a sense of ‘cause-to-move (by-blowing)’, whenever the verb occurs with the Caused-Motion syntactic argument structure [NP V NP PP]. The extended semantic structure of *blow* as the main verb of sentence 15 can then be translated directly into Hebrew to generate the main verb in the Hebrew target sentence.

- (15) English: The wind **blew** the boat off course
Hebrew: Haruax **hesita**(n.s.t-*hif'il*) et hasfina mimaslula.
 'The wind **shifted** the boat off its course'

There are two problems with this approach. One is noted by Goldberg (1995). Goldberg points out that the theory of Levin and others (e.g., Levin, 1985; Levin & Rappaport, 1988; and also Pinker, 1989) is forced to claim that a verb such as *blow* in English has several different senses (*blow*₁, *blow*₂, ...), one for each use (as in *the wind blew hard* vs. *the wind blew the ship off course*). However, the only evidence for the different senses of the verb *blow* is the fact that it occurs in the particular linguistic configurations given above. The motivation for assigning different senses to the verb is thus circular (and suggests that we may be dealing with ad-hoc assignment of senses)²⁶.

The problem with the lexical sense extension approach is even more acute when considering the translation of English caused-motion sentences into Hebrew. An underlying motivation in the “lexical sense-extension” line of research is the preservation of *compositionality* in semantics and in translation (i.e., that the composition of "meaning components" in the source text are translated into the target language). But note that for

²⁶ Goldberg suggests that the different semantics associated with the full clause when the verbs occurs in different syntactic environments should be attributed to the syntactic constructions themselves rather than to the verb (see presentation in section 1.2.1).

translation purposes it is not enough to extend the semantics of the verb ‘blow’, when occurring with the Caused-Motion argument structure, into a general caused-motion sense (i.e., ‘cause-to-move-by-blowing’). To preserve compositionality in translation, a semantic component identifying the particular type of motion event has to be incorporated into the extended semantic structure of *blow* as well (as translation into Hebrew suggests, see examples in section 8.4.1). The problem is again that the particular type of motion involved cannot be predicted in advance in the pre-encoded sense extension rule for the verb *blow*. It is the particular participants involved in every instance of the caused-motion macro-event that define the type of effected motion. The same verb *blow* or *sneeze* occurring with the *same* syntactic structure but with *different* participants instantiating the construction will evoke different types (manners) of motion (see examples 10-11 in the previous section).

9.3 Conclusions

Melby (1995:48) cites Minsky (1994) who said that one thing which separates current machines from humans is the *flexibility* of the human mind. When a computer program encounters a situation for which it has not been explicitly programmed, it either stops or produces meaningless results. When humans encounter a new situation, they are often able to try various solutions until something works. This description fits well the problem of processing novel linguistic blends: creative blends cannot be programmed in advance. However, when people encounter a new blend, they are usually able to reconstruct a possible set of correspondences (a mapping pattern) between the linguistic structure and a probable sequence of events in the world.

The goal for future research in NLP, I believe, is to take the notion of blending and linguistic creativity seriously, and conduct basic research to look for ways to computationally simulate creative blending processes, at least to some extent. The analysis in this chapter suggests that translation (of even very basic sentence structures) could never

be done accurately enough without incorporation of dynamic cognitive processes such as mapping, blending and integration of representational structures. So far, we may have identified only a small part of what is needed to make the modeling of dynamic general language processing possible, but this is a jumping off point for beginning to build machines that act more like humans do²⁷.

The discussion in this chapter also points to at least two reasons why MT systems in their current form can still produce partially successful results without completely simulating creative blending operations. One reason is that in spite of the immense potential for flexibility in the generation of linguistic expressions, much of language use is entrenched and predictable (if not in a deterministic way, then at least statistically), as suggested by the relative success of statistical NLP systems (section 8.1.2.1). The second reason is that in translation, it is very often the case that exactly the same linguistic blend (i.e., a similar integrating construction and a similar mapping pattern) is favored in both the source and the target languages to express an event. In such cases, the additional semantic structure imposed on the linguistic blends by human readers is transferred into the target language without being explicitly expressed in the translation.

In practical terms, the analysis in this chapter makes several points:

(1) That it is a mistake to try and account for every failure of an NLP system to provide a correct “model” (interpretation) for an input sentence by assuming necessarily that the *permanent* knowledge structures in the system are inaccurate and should be modified or extended. Often the knowledge structures are accurate and complete. The failure of the system may result from the speaker’s creative integration (blending) of these permanent

²⁷ I do not intend to claim here that the particular cognitive skills that humans use in language processing (whatever these skills are) are necessarily the only right ones for an NLP machine to use. However, I do suggest that NLP systems cannot afford to ignore these type of processes and skills and the cognitive power they provide.

knowledge structures into new temporary structures.

(2) Pre-encoded inference rules can capture only the most entrenched (repeating) instances of blending. They cannot solve the core problem of blending. Reconstruction of blends has to be performed on-line, imitating human cognitive creativity in finding possible correspondences between linguistic forms and complex events in the world. Finding correspondences involves abstracting previous encountered events encoded in memory, and searching for optimal mappings between fragments of retrieved structures and the information communicated in the linguistic utterance, in a way which complies with the grammatical blending conventions of the language²⁸.

Even if blending mechanisms cannot be completely automated with current computational techniques, it is still important to realize which aspects of failures of NLP systems are due simply to a *scale* problem (which more powerful computers and better algorithms can solve), and which are due to the very *nature* of language processing versus current computational techniques .

(3) Computationally, the mechanisms of language understanding discussed in this chapter (such as the de-integration mechanisms) are very different from the traditional AI logical inference mechanisms. In general, while logical inference rules necessarily have a single definite outcome, the mechanisms of language interpretation discussed in this dissertation define only general *procedures* triggered by various grammatical forms (e.g., a grammatical form such as the English Caused-Motion construction, or the Hebrew *binyanim*, trigger “de-integration” procedures guided by specific constraints), whose actual

²⁸ Note that the basic problem of *perceiving analogies* incorporates in itself the problem of simulating cognitive creativity. Douglas Hofstadter (1995b) points out that most current computational models of analogy-making erroneously incorporate the structural similarity into the structure of the input domains beforehand. This form of modeling fails to capture the *creative* aspect of analogy-making where patterns of similarity dynamically alter from one context to another. Some models of analogy-making such as "Copycat" (Mitchell and Hofstadter, 1990, 1993) incorporate some level of dynamic on-line similarity detection that changes in context (see also the discussion by Holyoak and Barnden, 1994, where they point that "Copycat" is guided by "*soft* pressures rather than *rigid* requirements", as in rule-based AI systems).

application varies based on the semantic and ontological properties associated with the particular linguistic instance. In other words, while traditional lexical inference rules (e.g., sense extension rules) are in the general form of “if X has property ‘p’ (in context Y) then add (or replace) property ‘q’ in X”, the mechanisms discussed in this thesis are of the general form of “if X has property ‘p’ (in context Y) then *apply procedure q* to X”. The outcome of applying procedure q to X depends on the semantic properties of X and its context Y, and thus varies for different instances of X. What guides and constrains these procedures requires better understanding of human linguistic blending processes. A final note should be made about the implications of the discussion in this chapter for research on *human-aided* Machine Translation. Many recent MT projects have dropped the requirement that the MT system would be fully automatic, and include instead some form of human-machine interaction during the process of translation, typically with the computer asking questions and the human partner answering them. Currently no one has come close to a successful interaction between human and machine in translation. Particularly, it is not yet clear what kind of questions should be posed by the system to the human, at what stages of the translation process, and how would the interaction proceed. The essential idea is for the NLP system to automatically process the “low-level” simple tasks, but to interact with the human user in making the more difficult decisions (e.g., syntactic and semantic disambiguation (Nirenburg *et al.*, 1992)). Depending on the approach taken to language processing and translation, the type of human-machine interaction changes drastically.

The analysis of translation examples in this manuscript suggests that an integral part of the translation process is the inferring of *additional information* beyond what is explicitly provided in the text (i.e., beyond, for example, the disambiguation of explicit linguistic information). The discussion in this chapter also suggests that this task is currently beyond the computational power of NLP, since the extra information often cannot be derived logically by pre-encoded inference rules, but rather involves novel complex manipulation of

existing knowledge structures (through processes of analogy making, mapping, and integration of partial structures). However, in contrast to NLP systems, humans seem to excel in these types of processes (as evident in the mostly flawless processing of creative linguistic blends by humans). The straightforward conclusion, is that the part of the translation process, where extra information must be added to the semantic representation of a text to enable its correct translation, is especially suited for human intervention.

Pierre Isabelle (1993) notes that within the ‘human-aided MT’ paradigm, “humans have persistently been asked to do things they would rather not do,, like answering odd questions about phrase bracketing or rearranging bizarre jumbles of target language words” (p. 202). Rather than tiring human partners with unintelligible machine-like tasks, we should take advantage of what humans are best at, and what machines are worst at: making cognitive decisions based on detailed knowledge of prototypical events (episodic memory). In such a system, it would be the task of the MT system to *identify* the need to augment a semantic representation of a text, but it would be the human partner who would actually provide the additional semantic-pragmatic content (when prompted by the machine).

For example, consider the translation of English caused-motion sentences into Hebrew or French discussed in chapters 8-9. In symbolic interlingual MT, the system could identify (based on the syntactic form of the English input sentence, and the novel use of the verb in the construction) that the input sentence communicates a novel linguistic integration of a caused-motion event sequence. The MT system would then prompt the human partner for help. A sophisticated interactive system would also present the human participant with information extracted from the linguistic structure based on pre-encoded rules (e.g., the system would identify the agent, moving patient, and the direction of motion in the communicated caused-motion event). The system would then ask the human partner to identify a probable “causing” and/or “effected” motion event. The information required from the human partner could also be guided by information on the type of integrating

linguistic constructions available in the target language, and the particular aspects of events most commonly encoded via the target language constructions.