A Survey and Critique of

American Sign Language Natural Language Generation and Machine Translation Systems

Technical Report
Computer and Information Sciences
University of Pennsylvania

MS-CIS-03-32

Matthew P. Huenerfauth

September 2003

Table of Contents

Abstract

This report contains a comparison and analysis of four of the most promising research systems for the translation of English text into American Sign Language animations. Beginning with an overview of ASL linguistics, a discussion of the special challenges of a language without a writing system, an explanation of the use of human figure animations, and a motivation for this MT task, the report continues on to introduce the four systems under consideration: the ViSiCAST translator [Marshall & Sáfár 2001], the ZARDOZ system [Veale et al. 1998], the ASL Workbench [Speers 2001], and the TEAM system [Zhao et al. 2000]. The systems are compared in terms of their MT architecture, grammar formalisms, linguistic representations, lexicon format, grammatical coverage, handling of classifier predicates, development time, and other factors. Strengths of individual systems are identified in the areas of spatial reasoning, modeling of ASL phonology, discourse representation, and motion parameterization.

A Survey and Critique of

American Sign Language Natural Language Generation and Machine Translation Systems

Matt Huenerfauth
Technical Report MS-CIS-03-32
Computer and Information Sciences
University of Pennsylvania
September 2003

The Problem of ASL Generation

This report is a survey and critique of several of the most successful English to American Sign Language (ASL) machine translation systems. This report will first introduce the reader to the special properties of ASL that make it a challenging subject of computational linguistic study and will then frame the problem of English to ASL translation. After the four systems in this survey are introduced, they will be compared in a thematic fashion – according to several criteria important to a successful English-to-ASL translation system. Finally, each of the systems will be discussed in terms of a particular strength or interesting approach it has taken to the ASL MT task.

The Linguistics of American Sign Language

American Sign Language is a natural visual/spatial language used primarily by a half a million deaf individuals in the United States and Canada. Modern linguistic analysis of American Sign Language is still in its infancy compared to that of other natural human languages. Up until quite recently, there was strong debate over ASL's status as a language, and initial analyses of its structure suffered from English biases, difficulty in accurate data collection, and failure to appreciate its full manual and non-manual modes of expression [Neidle et al. 2000]. Some of the initial claims from this

early research were that ASL was not a true language, possessed free word order, or had a flat non-hierarchical syntactic structure. Modern linguistic analyses have disproved these claims and have justified its status as a true natural language by distinguishing it from a simple gestural system or from a manually signed form of English.

This modern research has shown that ASL has a distinct grammar, vocabulary, and structure from English and that its visual nature allows it to use linguistic phenomena not seen in any spoken language [Neidle et al. 2000] [Valli & Lucas 2000]. As opposed to traditional natural languages which use spoken sounds or written symbols, ASL relies on the multiple channels of handshapes, movements, facial expressions, and other non-manual signals (NMS) to convey its meaning. Traditional languages typically lengthen a sentence by appending morphemes or adding lexical items in order to incorporate additional information, but ASL often makes use of its many channels to incorporate additional information by modifying the performance of a sign, performing a meaningful facial expression during a sentence, or making use of the space around the signer.

ASL signers will use the space around them for several grammatical, discourse, and descriptive purposes. During a conversation, an entity under discussion can be "positioned" in an imaginary point in space around the signer. Subsequent pronominal references to this entity can be made by pointing to this location, and many forms of verb agreement will use this point in space to help indicate the subject, object, or both. Signers can also use the space around them in a more three-dimensional pantomimic fashion. Some verbs can incorporate a movement path into their performance that indicates information about the manner or direction of activities in the real world. Special ASL constructions called "classifier predicates" allow signers to use their hands to represent an object in the space in front of them and to position, move, trace, or re-orient this imaginary object in order to indicate the movement of the object under discussion. This spatial and three-dimensional nature of ASL is a particular challenge to the use of traditional computational linguistic approaches to the ASL translation task.

The Concept of Animated Avatars

Research into virtual reality human modeling and animation has reached the point of sophistication where it is now possible to construct a model of the human form which

is articulate and responsive enough to perform American Sign Language. The level of quality of such human avatar animations has increased such that human signers of ASL can now view the onscreen animations and successfully interpret the movements of the avatar to understand its meaning [Wideman & Sims 1998]. However, just because graphics researchers know how to move the character, doesn't mean that we have ASL generators. Given an English text or abstract semantic input, a computational linguistic component would need to tell the animated character what to do (assuming the correct instruction set for the interface between the linguistics and the animation components is already determined). The English-to-ASL translation systems in this survey take an English text input and produce this ASL output "script" for an animated avatar to follow. Because ASL is a language without a standard writing system, the form of this script specification varies between the surveyed systems and is an open area of debate.

A Language without a Writing System

When considering generation of American Sign Language, it is natural to consult research in generation for traditional written/spoken languages. However, because of the unique complexities of signed languages, the generation problem must be framed slightly differently. ASL does not have a standard written notation accepted by the deaf community; so, there is no form of "text" which can be produced by the generation component. Instead, it is useful to think of the production of an ASL animation as analogous to the entire semantics-to-speech generation process for a traditional language. The ASL animation via an onscreen avatar shares many of the qualities of audio output speech; it is ephemeral, performed over time, and can represent subtle linguistic and emotive variations via prosodic or performative modulations to the standard form. This lack of a standard written notation for ASL is one of the reasons why collecting transcribed corpora of the language is extremely difficult and expensive, and this lack of corpora is why none of the surveyed systems attempt a statistical approach to ASL MT.

To help modularize the ASL production process and to bring the ASL generation problem closer to that of traditional written/spoken languages, many researchers (including several of those in this survey) have proposed written representation systems for ASL. All such abstractions of an actual ASL performance will omit some amount of

detail, and choosing what detail is acceptable to omit when developing an artificial writing system for a natural language is a challenging and error-prone task. We shall see that those systems whose notation system for ASL is insufficiently expressive will limit their ability to produce a fluent signing performance.

Motivations and Applications

Building a generation system for American Sign Language is important because although deaf students in the U.S. and Canada are taught written English, the difficulties in acquiring a spoken language for students with hearing impairments results in the majority of deaf high school graduates in the U.S. having only a fourth-grade reading level [Holt 1991]. Unfortunately, many of the approaches taken to making elements of the hearing world accessible to deaf individuals (such as television closed captioning or teletype telephone services) assume that the viewer has strong English literacy skills. Since many of these individuals may have full linguistic fluency in ASL despite their difficulties with written English, an automated ASL translator could make more information and services accessible in situations where the English captioning is beyond the reading level of the viewer or a live interpreter would be inappropriate.

An automatic English-to-ASL translation system could also enable several new educational applications for deaf students learning English or hearing students learning ASL. English reading educational software that could explain a text in sign language at the press of a button would be a great resource for English teachers at Schools for the Deaf. Hearing parents of deaf children or others interested in learning ASL could also benefit from a language educational program which could teach them particular sentences or phrases on demand when they type in an English source text.

* * *

Brief Introduction to the Systems Under Consideration

This paper will focus on the four sign language translation projects that come closest to the goal of automatic translation from English input text to fluent American Sign Language animation output. There are many systems which develop only subcomponents of a sign translation tool or which only attempt to translate from English

to a manually represented form of English, like Signed Exact English. These projects gloss over the linguistic divergences between English and American Sign Language, and therefore they are not included in this study. The next section of this report will survey these four systems, and it will critique and compare them in an issue-by-issue basis. However, this section will first briefly introduce each system to help the reader differentiate them in the comparative discussion that follows.

ViSiCAST Translator

As part of the European Union's ViSiCAST project, Ian Marshall and Éva Sáfár at the University of East Anglia implemented a system for translating from English text into British Sign Language [Marshall & Sáfár 2001] [Sáfár & Marshall 2000] [Sáfár & Marshall 2002] [Bangham et al. 2000]. The system was also considered a research vehicle for translation to German or Dutch sign languages, but that capability has not yet been implemented. Their approach uses the CMU Link Parser to analyze an input English text, and then uses Prolog declarative clause grammar rules to convert this linkage output into a Discourse Representation Structure (DRS). During the generation half of the translation process, Head Driven Phrase Structure rules are used to produce a symbolic sign language representation script. This script is in the system's proprietary "Signing Gesture Markup Language," a symbolic coding scheme for the movements required to perform a natural sign language [Kennaway 2001].

ZARDOZ

The ZARDOZ system [Veale et al. 1998] was a proposed (and somewhat prototyped) English-to-Sign-Languages translation system using a set of hand-coded schemata as an interlingua for a translation component. (The discussion herein will focus on the ambitious proposed architecture rather than the implemented portion.) While the implemented portion of the system focused on American Sign Language, the authors were developing their framework with British, Irish, and Japanese Sign Language also in mind. Some of the research foci of this system were the use of AI knowledge representation, metaphorical reasoning, and a blackboard system architecture; so, the translation design is very knowledge and reasoning heavy. During the analysis stage, English text would undergo sophisticated idiomatic concept decomposition before

syntactic parsing in order to fill slots of particular concept/event/situation schemata. The advantage of the logical propositions and labeled slots provided by a schemata-architecture was that commonsense and other reasoning components in the system could later easily operate on the semantic information. Because of the amount of handcoding needed to produce new schemata, the system would only be feasible for limited domains. To compensate for these limitations, if a schema did not exist for a particular input text, the system would instead perform sign-for-word transliteration.

ASL Workbench

D'Armond Speers proposed and implemented an ASL machine translation system called the ASL Workbench [Speers 2001], which is based on a thorough understanding of modern ASL linguistic research. His approach uses a lexical-functional grammar (LFG) for analysis of the English text into a functional structure, hand-crafted transfer rules for converting an English f-structure into an ASL one, and LFG rules to produce ASL output. The system uses a transfer-specific lexicon to map English words/phrases to analogous ASL signs/phrases. When the system encounters difficulties in lexical choice or other translation tasks, it asks the user of the system for advice. The system also produces a very simplistic discourse model from the English input (consisting of a flat list of discourse elements and their spatial locations if yet specified), but all reference resolution must be performed manually by the human operator. The sophisticated phonological model used by this system is particularly robust and is based on the modern Movement-Hold model of ASL phonology [Liddell & Johnson 1989].

TEAM Project

TEAM was an English-to-ASL translation system built at the University of Pennsylvania that employed Synchronous Tree Adjoining Grammar rules to build an ASL syntactic structure while an English dependency tree was built during analysis [Zhao et al. 2000]. The output of the linguistic portion of the system was a written ASL "gloss notation with embedded parameters" that encoded limited information about morphological variations, facial expressions, and sentence mood. This project took advantage of the virtual human modeling research also being performed at the University of Pennsylvania by using one of the Human Modeling and Simulation laboratory's

animated virtual humans as the signing avatar. While the extensibility of the system's STAG translation approach can be questioned because of the simplifying assumptions inherent to the system's "gloss" notation, the project had particular success at generating aspectual and adverbial information in ASL using the emotive capabilities of the animated character.

* * *

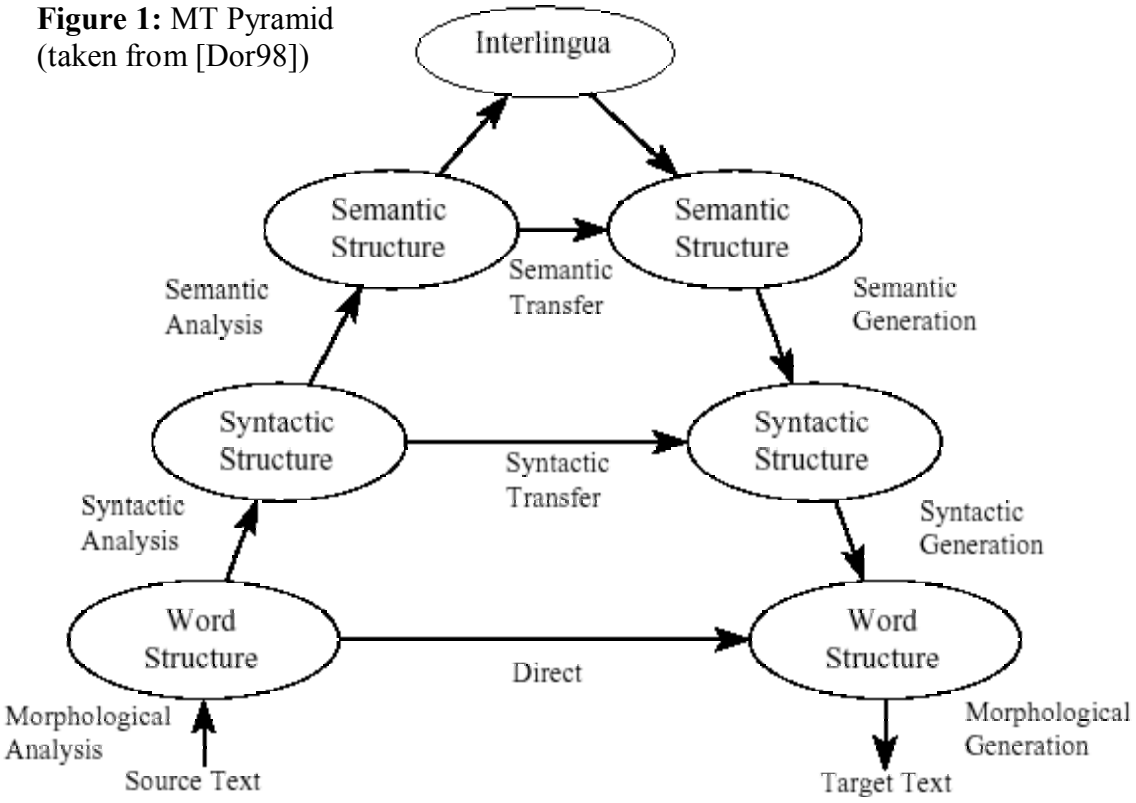
Thematic Critique and Comparison of the Systems

This section of the report will highlight several issues or design features that are important to consider when evaluating the success of an English-to-ASL machine translation system. As each topic is introduced, the systems in the survey will be analyzed and compared according to the criterion. The issues discussed range from architectural concerns such as the MT approach or grammar formalism to practical considerations over the development time invested in each system or the degree of user interaction required for the translation system to function correctly.

Underlying Machine Translation Architecture

There is a spectrum of architectural designs along which most machine translation systems can be classified, and loosely they are grouped into three basic designs: direct, transfer, or interlingua [Dorr et al. 1998]. Direct systems base their processing on the individual words of the source language string; translation is achieved without performing any form of syntactic analysis on the original input text. Transfer systems do analyze the input text to some syntactic or semantic level, and then a special set of "transfer" rules are employed to read the information of the source language structure and produce a corresponding syntactic or semantic structure in the target language. Afterwards, a generation component is used to convert this linguistic structure into a target-language surface form. Interlingual systems take this analysis of the input text one step further: the source is analyzed and semantically processed to produce a typically language-independent semantic representation structure called an interlingua, and then a generation component produces the target-language surface form from this starting point. These three architectures are commonly visualized as a pyramid as in Figure 1.

Figure 1: MT Pyramid (taken from [Dor98])



Generally, in the absence of statistical or case-based information, the higher up along the pyramid that the source text is analyzed, the more complex and subtle are the divergences that the system can handle. In particular, as one reaches the level of interlingua, generally a knowledge base is used to supplement the linguistic information, producing translations that use information about the world and that may convey more information than was present in the source text (devoid of context). However, any of the approaches can produce a correct translation for certain inputs. Another trend as one goes up the MT pyramid is that the amount of domain specific development work that must be performed increases dramatically. (Constructing interlingua representations and knowledge bases for all possible translation situations is daunting if not impossible.)

The reason for the caveat in the previous paragraph excluding systems that use statistical or case-based information from these MT Pyramid generalizations is that such data can be used to make the development work vs. divergence handling power trade-off less severe. Empirical information can inform the decision making process during translation for each of the architectures (direct, transfer, interlingua) described above. This information could be taken from successful examples of English to ASL translation. Since lexical, structural, semantic, and general-knowledge information may have been used by humans during the translation of these examples, then all such forms of information are implicitly captured in the example or statistics of both approaches. So, when this information is applied to a direct or transfer architecture, in effect, the system is gaining information from all of the possible levels of analysis, understanding, and generation. Unfortunately, the difficulty of collecting and annotating corpora for a signed language means that MT designers for ASL do not have a source of empirical data for producing a statistical system.

Some of the earliest attempts at English to sign language translation used a "direct" architecture; they attempted to structure the translation task as a word-to-sign replacement process. A lexicon of ASL sign animations was produced (from animating a virtual character, from using motion-capture glove technology, or from actual video recordings), and each entry in this lexicon was associated with a set of corresponding English words that it could be used to represent. This lexicon-building approach is consistent with one English-to-Sign translation system that was not included in this

survey: the TESSA system [Bangham et al. 2000]. TESSA takes an English input text, looks up each word of the English string in the English-to-Sign dictionary, concatenates those signs together, and blends them into an animation. Since such a system does not represent any of the underlying linguistic structures of English or ASL, the translation divergences between the two are difficult to handle. TESSA used a small set of standard sign sentence templates to compensate for some of these phenomena, but the use of templates in this manner is not scalable. For this reason, the system could never translate from English into a natural sign language, and instead merely encodes an English string visually as Sign-Supported English (a signing system which uses the exact grammar of English and adds signed hand motions to accompany content words). Because this system did not attempt to translate into ASL, it was not included in the survey.

Another lexical-based translation architecture for ASL was the TEAM [Zhao et al. 2000] system. This project used a synchronous lexicalized TAG grammar for English and for ASL, where a bilingual lexicon stored English word / ASL sign pairs. Each of the members of the pair stored a piece of TAG structure that represented the syntactic environment surrounding the lexical item in which that particular bilingual word/sign pairing would be valid. As the English input text was analyzed with a TAG parser, the English lexical entries were looked up, and the ASL signs to which they corresponded, identified. Simultaneous to the TAG analysis, the TAG generator for ASL worked to assemble the ASL TAG trees into a complete sentence. Because of the increased linguistic information, this architecture could handle many syntactic divergences between English and ASL.

While the TEAM approach may sound like a direct architecture since there seems to be a word-to-sign mapping, it is actually a syntactic transfer approach. The English input string needs to be analyzed with the TAG parser during the translation process, and the syntactic information revealed helps direct the bilingual lexicon look-up process. The "transfer rules" in this system would be each of the paired entries in the bilingual lexicon; by identifying and applying this matching process, we convert a syntactic analysis of the English sentence into a syntactic structure for ASL.

Another English-to-ASL machine translation tool using a "transfer" architecture was the ASL Workbench system [Speers 2001]. This system performed deeper linguistic analysis than the TEAM system above; instead of stopping at a basic syntactic constituent structure, this system analyzed the input English text up to an LFG-style f-structure. This representation abstracts away some of the syntactic specifics of the input text and replaces them with linguistic features of the text, such as "passive voice," or functional labels on constituents, like "subject" or "direct object." This Workbench system included a set of specially written rules for translating an f-structure representation of English into one for ASL.

While this system's deeper level of analysis may help it handle some more subtle forms of divergence than a system with more shallow analysis of the English source, the biggest advantage of analysis up to f-structure is that it simplifies writing transfer rules. Instead of conditioning on constituent structures of a syntactic tree, the rules reason about abstract features or functional role labels in the English f-structure. Another advantage of deeper transfer is that generation for the ASL side begins with a more abstract representation. While this may sound like more work for the ASL generator, it actually gives the system more flexibility in its choice of structure. For example, although Workbench only handled NMS minimally, crossing at f-structure could make it easier to select if, where, and how to express NMS over the sentence. One problem with the TEAM system was that it tried to produce ASL NMS features from looking at the English syntax tree – an inappropriate method for determining the correct NMS for an ASL syntactic sentence being built. Making the transfer crossing at a deeper level of representation could have addressed this issue.

ViSiCAST is another system with a transfer-based architecture; in this case, the cross-over from the English analysis to the ASL generation occurs at a semantic level of representation. The English portion of the system converts a text into a set of variables and semantic predicates in a Discourse Representation Structure. After a conversion process to a hierarchical semantic format, an HPSG generation system begins the ASL production process. The types of divergences that can be handled by such a system are more sophisticated than those that the syntactic transfer TEAM or Workbench systems can accommodate; the use of semantic predicates helps to break apart the individual

elements of meaning in the English source text and helps avoid an English syntactic influence on the ASL being generated. However, these additional levels of representation and syntactic-semantic linkages significantly increase the development time of this system compared to those approaches that do not go as high up the machine translation pyramid.

The ZARDOZ system falls into the final category of English-to-ASL translation systems – those which use an interlingual representation. The authors of this system identified many types of semantic divergences between English and ASL which could only be successfully handled by a system that would be able to employ some form of spatial or commonsense reasoning. (Most of these issues surround the use of classifier predicates.) So, in their proposal, the input English text would be analyzed syntactically and semantically, and this information would be used to select a particular event schema. These schemata would record all the common types of events and actions occurring in the world (and their participants); so, the time needed to develop this set of schemata restricts this system to a small domain. If a schema were filled using the information from the English string and any world knowledge, then a generation component could be used to express the information stored in the schema in ASL animation output. Because there was a common representation structure used for both English and ASL, and because these schemata were generally language-independent, they can be considered an interlingua.

Coverage, Specificity, and Development Time

The four systems in this survey vary widely in the amount of grammar and lexicon development time that has been devoted to them. Of the four, the ViSiCAST project is the most robust, and it is the only system still under development and with recent publications. While it only has a lexicon of approximately 50 entries, the designers claim that the entries in this lexicon were chosen for their linguistic breadth. Their research focus has been on incorporating progressively more complex linguistic phenomena into the system's grammar. The system's ASL grammar can currently process approximately 50% of the CMU English parser's "link types." For example, the system can build ASL structures from English ones involving "transitive and intransitive verbs, temporal auxiliaries, passive, imperative," subject/object relative clauses, "determiners, polite

requests, expletives, predicatives, pronouns, wh-questions, yes-no questions, and negation." [Marshall and Sáfár 2001]

In contrast, both the TEAM and Workbench systems were short-lived research projects that used small lexicons and grammars in order to demonstrate the potential of their approaches, not to provide broad linguistic coverage. The TEAM system's small sign vocabulary and limited number of ASL STAG trees meant that there was a limited number of demonstration sentences that the system was able to successfully translate. While the TEAM system is no longer under development, there is work in ASL machine translation at the University of Pennsylvania that has grown out of this project [Huenerfauth 2003]. The ASL Workbench was a brief linguistics dissertation project, and it, too, is also no longer under development. While the author included a discussion of significant structural divergences between ASL and English, these analyses were not reflected in the system's implemented transfer correspondence rules. The system's ASL generation grammar was more robust; the author surveyed several ASL syntactic constructions and formalized them in LFG rules.

While it is the development of grammar, lexicon, or transfer rules that limit the coverage of the previous three systems, for ZARDOZ, the creation of interlingual schemata for each new domain would be the most significant bottleneck to development. The system does allow a form of "back-off" that somewhat diminishes this domain specificity; however, the reader should note that what the authors mean by back-off is different from the traditional MT or statistical sense. When the interlingual resources of ZARDOZ do not cover a particular word in the input English text, then the system does not attempt a different form of translation, instead it merely attempts a word-to-sign mapping from English into another coding system for English, Signed Exact English. This transformation is a very simple rule-governed process, not another form of translation. In fact, the authors propose mixing ASL and SEE output within a single clause or sentence; while such mixed output might possibly be understandable to a signer, it would not be fluent. Of course, since the ZARDOZ system was only minimally implemented, many of the details of this backing-off process suggested by the authors were never fully developed, and further, this research project is no longer active. In fact, the practicality or degree of implementation of the generation approach was not a primary

motivator for the designers of this system; these authors were primarily interested in developing a vehicle for their artificial intelligence research work in metaphorical reasoning, knowledge representation, and blackboard architectures.

ASL Grammar Formalism and Generation Approach

The ViSiCAST system employed Head Driven Phrase Structure rules for the generation component; this grammar formalism is well suited to ASL because it operates on multi-level feature structures, allowing access and modification of various levels of linguistic specification easily. In ASL in particular, a single piece of information can often be expressed at a variety of linguistic levels or modalities (via lexical choice, morphology, syntactic movement, timing of performance, manner of articulation, non-manual signals, etc), and it is useful to have access to all of them simultaneously when choosing a mode of expression.

While a feature-structure-modifying grammar would be a good choice, the phrase structure aspect of the formalism may be less suitable to ASL. Phrase structure-based grammar formalisms are particularly well-suited to languages in which much of the generation process is devoted to determining the word order of the final linguistic output. In ASL, the concept of a single word unit is harder to define, and the plethora of movement-producing and constituent deleting phenomena can make grammars that focus too much on building surface word order cumbersome to implement – there would be too many rules needed to capture all permutations resulting from movement and deletion. A system that decoupled the numerous movement and deletion processes of American Sign Language from the rest of the generation process and applied them at a later stage could be a more logical choice. Of the many interacting constraints to be considered during the ASL generation process, the word order of the manual signs in the sentence is actually one of the more flexible. Unfortunately, most phrase structure based generation approaches may narrow in on a particular word ordering too quickly and rigidly before making other generation choices.

During ASL generation in the Workbench system, the Lexico-Functional Grammar phrase structure rules first produce a syntactic frame from the functional representation and then create a stream of ASL movement/hold segments as the phonological output.

Under this particular architecture, all syntactic and lexical choice is made prior to morphological, phonological, or non-manual expression decisions. Considering the interactions and constraints that exist between these linguistic levels and the fact that some information in ASL (such as agreement) could optionally be expressed in several different levels or modalities, not considering morphological and non-manual capabilities at the same time as syntactic and lexical choice seems to be a mistake. In addition to this choice-ordering issue, the phrase structure generation approach on Workbench may suffer from similar word-ordering related rule count issues as discussed with the ViSiCAST system above.

While STAG grammars were successfully used to produce a functioning prototype in the TEAM system, part of their success can be attributed to simplifying assumptions made about ASL: the motivations for movement, the use of non-manual signals, the phonological complexity of signs, and the ability to fully represent ASL via a string of glosses. (Some of these assumptions may have been triggered by this project's reliance on older linguistic analyses of ASL.) Modern research has shown that much of the movement, topicalization, tagging, and referential choice in ASL is motivated by discourse-related factors not modeled in this translation system. From simply analyzing the tree structure of the English input, the surface tree structure of the ASL cannot be immediately determined – however, this is exactly what the TEAM system attempted to do. If the STAG system in this project had implemented a discourse model and then added a tree-revision and movement stage to its pipeline to modify the initial tree after the transfer process was complete, then some of these linguistic phenomena could be handled. The addition of a discourse model to the TEAM system could also facilitate the successful generation of pronouns and determiners that require knowledge of the location in space associated with a particular discourse entity. Certain ASL verbs also require agreement information to be expressed, but without a model of where the verb's subject or object exist in the discourse space around the signer, the system would not know what location parameters to pass to this agreement morphological component.

During the generation process of ZARDOZ, interlingual schemata are converted into symbolic ASL output; however, traditional syntax trees are not used during generation. Instead, a partial ordering graph structure called a spatial dependency graph

is incrementally built which can express partial ordering and simultaneity relationships between “case types drawn from a syntactic/semantic case ontology, indicating which elements are to be selected from the interlingua structure” [Veale et al. 1998]. Eventually this ordering information can be used to linearize the ASL into an output stream. Movement and other re-ordering/restructuring phenomena can be triggered by running special "style feature" detection routines on the original source language input, and when a particular feature is present, then the system would perform a manipulation operation on the graph structure. For example, if a "style" detector was run on an input English text, and it was determined that the object of the sentence was currently the topic of the discourse, then this may trigger a "topic feature" detector. This may trigger the ASL graph being built to rearrange itself so that the object of the sentence is moved to the front and is surrounded by topicalization NMS. This partial ordering graph-based approach could address some of phrase structure rule challenges of the ViSiCAST and Workbench systems above. The ordering constraints imposed by this formalism are partial and soft; so, the final manual sign order is left flexible as the generation process progresses.

Expressiveness of ASL Representations

The TEAM system’s use of “glosses with embedded parameters” [Zhao et al. 2000] to represent the ASL sentence being produced limits the quality of this system's non-manual signals and phonological smoothing. While some facial expressions are added to the glosses notation, they are indicated by inserting “begin-of-specific-nms” and “end-of-specific-nms” markers into the gloss string. Since all of the special motion parameterizations must be attached or embedded to specific lexical entries in the output string, it is necessary to insert these dummy NMS glosses in this fashion. This representation also fails to model the changing degree of NMS over time and the oftentimes optional spreading of NMS over entire constituents or just feature heads.

The use of a gloss notation tends to treat the location, handshape, orientation, and movement features of individual signs as discrete and unrelated to those of adjacent glosses. A human-model graphical smoothing approach was used to make the figure's motion non-jerky using the Parallel Transition Networks animation script; however, ASL actually uses a series of complex phonological processes to govern morphological

expression and smoothing. These processes often operate on movement/hold segments of signs and would be difficult to implement under the sign-discrete gloss notation of this approach.

The ViSiCAST system does a somewhat better job of representing its individual sign information. The lexicon stores phonological information about signs using the Signing Gesture Markup Language [Kennaway 2001]. Some signs are fully specified in their description and are considered "fixed": meaning they will not undergo phonological changes during realization. Other signs have an incomplete phonological specification that can be modified and determined during the generation process. This allows the ViSiCAST approach more flexibility than TEAM in blending the movements of adjacent signs. Unfortunately, the system failed to take advantage of this capability; it did not include bits associated with signs to indicate what types of morphological or phonological operations could apply to it or how these operations would modify the internal structure of the sign. To produce an ASL component that can correctly inflect and modify signs during generation, such information would be critical.

ZARDOZ's representation of ASL is particularly insufficient in its treatment of ASL non-manual signals. At the realization level, the system inserts symbolic tokens into the output stream to indicate various atomic non-manual features (such as "eyebrows-downward"); a corresponding token is inserted later in the stream called "resume-previous-face." Even if atomic operators could be identified for all types of non-manual expression, this token-insertion approach is insufficient for representing the full spectrum of overlapped, simultaneous, and interacting non-manual forms of expression. The change in intensity of NMS over time (in particular in response to proximity to various feature heads in the output) also cannot be expressed. The architecture fails to correctly represent the optional nature of NMS expression and how this optionality is affected by the presence of other linguistic phenomena in the output. This token approach is quite similar to the “begin-specific-nms”/”end-specific-nms” approach taken by the TEAM system, but it is even less powerful. The use of a “resume” token as opposed to a specific ending token for each type of NMS means that the ZARDOZ system assumes that all onsets and completions of NMS features will have a nested parenthesis structure. While this appears to generally be the case for ASL, it may not be for more complex ASL

constructions or for all of the signed languages under consideration by the ZARDOZ project.

The ASL Workbench also does a poor job of representing the optionality, capability, and degree of expression of non-manual signals. An ASL signer has choices about whether to use NMS, over what extent of the sentence to express it, and to what degree of expression, but these decisions are based upon the lexical and syntactic choices made for the sentence and other physical limitations on the concurrency of NMS expression. Such NMS decision complexities are glossed over in this system, as is the ASL head-tile/eye-gaze verb agreement mechanism [Niedle et al. 2000] – a phenomenon responsible for much of this NMS planning complexity. Even if the system were enhanced such that the generation formalism incorporated more NMS decision capabilities, the otherwise rich phonological specification used in this project is very weak in its ability to model the NMS channel of output. While the phonological specification for the details of the manual movement is very detailed and precise, the Movement-Hold segments have only one slot per time-slice segment devoted to the NMS channel. Information about the simultaneity or degree of non-manual signals could not be captured in this formalism.

The Animation Output

Qualitative analysis of the animation output of an English-to-ASL translation system is also an important consideration when determining the success of a particular project. However, only one of the systems in the survey, TEAM, actually produced any animated American Sign Language output to consider. The ViSiCAST system was designed to produce British Sign Language, and the Workbench project never successfully implemented an animation component. The ZARDOZ system was primarily an ambitious architectural proposal that was minimally implemented, and the system only produced a very small number of demonstration animations. For this reason, a true comparison of the animation output quality is not possible between the four systems. Instead, this section will consider how design features of the systems impact how the animation output would appear.

As discussed in the introduction, ASL is a language without a written notation system, making it a particularly difficult candidate for computational linguistic

processing. While it is theoretically possible for a machine translation system to analyze an English text and to immediately make decisions about the appearance of an ASL animation, all practical approaches to the English-to-ASL problem instead attempt to first construct a symbolic representation script of the ASL to be performed. Without a standard notation available, most of these systems have invented their own proprietary script format. In fact, since the ASL Workbench never actually implemented the animation portion of the system, this symbolic ASL script is the final output of this system.

By adopting a written output form of ASL, there is risk that the scope of the generation process has been reduced. If the output form is too underspecified a representation of ASL, then there still may be linguistic work to be done that the generator has failed to accomplish. Unless the writing notation adopted by the generation system is able to fully specify all of the movements of the avatar such that the animation could be produced without introducing ambiguity, then such a system has successfully accomplish only a part of the generation task. In a system like ASL Workbench that does not produce an animation output, it is difficult to decide how well the symbolic output has successfully translated from the original English text. The critique of ASL representation approaches in the previous section has suggested some areas in which Workbench's representation may omit too much detail – particularly concerning non-manual signals. It is very likely that if connected to an animation component, this system would produce ASL with less than fluent NMS expression.

Considering a system with an implemented animation component, like the TEAM project, it is easier to see where problems arise in the animation output. In fact, the critique of TEAM's "gloss" notation has already suggested where problems arise in the output. Because its symbolic ASL formalism possesses only limited controls for some non-manual signals, the TEAM system fails to implement the head-tile/eye-gaze approach to agreement expression, nor does it successfully indicate how the degree of expression of a non-manual signal should vary according to its proximity to particular lexical items in the manual signing channel. The monolithic manner in which the system represents the animation movement details of individual ASL signs also prevents the

system from blending signs smoothly into one another in a manner that is consistent with the phonological processes of ASL.

The ViSiCAST project is located at the University of East Anglia (in the United Kingdom) and is funded in part by the European Union; therefore, the goal of the system is to translate from English text into British Sign Language (BSL). Differences between ASL and BSL prevent this technology from being immediately adoptable to an American context. BSL's standard method of expressing the pluralization of nouns and verbs is to repeat the sign or sweep during its final hold. In addition to expressing plural differently, BSL expresses number information more frequently than ASL; in this system, nouns and verbs are pluralized whenever possible. ASL can occasionally express plural though internal morphological changes or repetition of classifier predicates, but this is the exception, not the rule. Generally number is conveyed via the determiner, but its expression is optional – this optionality is not modeled in the ViSiCAST system.

Non-Manual and lexical differences would also make the ViSiCAST system difficult to use for ASL. The use of non-manual signals in BSL is much more restricted and infrequent than in ASL. For example, this project makes no effort to implement head-tilt/eye-gaze agreement, partially because it has not yet been shown that BSL or the other European sign languages employ it. The representation of NMS in the HamNoSys [Prillwitz et al. 1989] output notation is also quite limited, and while it would be insufficient for ASL, it may suffice for the understated NMS typical of BSL. While not yet implemented, the authors also discussed building a "mouthing" capability into the BSL system. Unlike ASL, British Sign Language is often accompanied by the signer mouthing English words to accompany various signs. Perhaps most trivially in terms of the system's architecture, the forms of fingerspelling and the vocabulary of the two languages are so different as to be mutually unintelligible, but these problems could easily be addressed by filling the lexicons and coding the fingerspellers differently.

Classifier Predicates and the Use of Three-Dimensional Space

In ASL, a classifier predicate is a complex (often spatially semantic) sign which typically follows a noun phrase to produce an ASL utterance. Classifier predicates challenge traditional definitions of what constitutes linguistic expression, and they

oftentimes incorporate spatial metaphor and scene-visualization to such a degree that they verge on pantomime or 3d-modelling. A classifier predicate is created by first selecting one of a closed set of handshapes based on the characteristics of the entity in the noun phrase (whether it be a vehicle, upright animate figure, squat four-legged object, etc.) and what aspect of the entity the signer wishes to discuss (its surface, size, position, motion, etc). The signer then invents a three-dimensional movement for the hand which communicates a contour, position in the space around the signer, motion through 3d-space, a physical/abstract dimension, or some other property of the object which needs to be communicated. Classifier predicates are therefore ideal for describing scenes, articulation of tools, movements, sizes, and other information of a visual/spatial or scene/process nature.

The complexity and variability of these signs has led some ASL machine translation researchers to omit classifier predicates from their systems: the ViSiCAST and TEAM projects do not make any effort to address classifier predicate constructions. However, the frequency of these predicates in ASL discourse and their elegant expressive capabilities makes these constructions a common and powerful communicative tool for ASL signers (and potentially for an ASL generator). While the Workbench system did not address the problem of how to produce a classifier predicate from an English source text, it did leave room for growth in its system design such that classifier predicates could be incorporated into the generation process. Speers chose to represent classifier roots as highly underspecified lexical entries whose movements would be determined by the generation grammar; thus they could be treated like any other single lexical item. He also categorized the handshapes and movement types of common classifier predicates.

While the ZARDOZ system was only a minimally implemented proposed system design, the proposal did include some ambitious ways in which classifier predicates could be generated by the interlingual architecture. In this design, particular motifs of classifier predicate expression could be represented by unique interlingual frames that could be selected and filled by the analysis/understanding portion of the translation architecture. The proposal discussed how spatial and commonsense reasoning approaches could be used to fill in the animation details needed to produce a fluent classifier predicate handshape and movement. Taking into account the state of the art of AI reasoning and

spatial representation technology, the development of such a system would presently be a very domain specific and time-consuming task. While the ZARDOZ approach is not a practical system design, there are elements of this approach in the Parameterized Action Representation (PAR) based translation design proposal of [Huenerfauth 2003]. Both systems use an interlingual approach to handling classifier predicates, require the entities in a scene to be represented, and require the specifics of the movement script to be filled in prior to animation time. The difference is that the latter approach grounds this representation and reasoning process in previously developed technologies of the Human Modeling and Simulation Lab, namely the virtual human and scene modeling system and the Parameterized Action Representation movement planning approach [Badler et al. 1998].

Sign Lexicon Specification

Each of the four systems in this survey takes a different approach to how the movement and animation details of a sign lexical item are recorded. The TEAM system used a very small demonstration lexicon, and the animation movements of the individual signs were stored as parameterized motion path templates compatible with the Jack Toolkit and Jack Visualizer [Badler et al. 1998]. In this manner, the movements were defined in a graphical fashion using the proprietary control language of Human Modeling and Simulation lab technology. The sophisticated movement and animation planning capabilities of this framework allowed for the flexible specification of an action or movement path with a small number of input parameters. For example, the system is capable of creating a 3-D motion path that intersects a set of desired "goal" and "via" locations.

Similarly, the ZARDOZ system defines its sign lexical items using code segments written in the Doll Control Language [Veale et al. 1998]. Again, a human avatar movement coding system is used as the basis for defining the animation details of the sign lexicon entries. While TEAM and ZARDOZ use a graphical control language of sorts, the TEAM system processes more abstract and parameterized motion commands than the Doll Control Language. Another difference is that the phonological movement specifications for ZARDOZ are stored in a hierarchical lexicon in which signs with

related performance features are grouped. Therefore, the full animation script for a particular sign is based on the inheritance of the movement specifications of its ancestors in the hierarchy. While the Parameterized Action Representation motion planning operators of the TEAM system are amenable to this same type of hierarchical specification approach, the small lexicon of the TEAM system did not take advantage of this capability.

The ViSiCAST system uses a more sign-language-specific scheme for defining the animation details for its lexical items. The ASL animation script processed by this system is the "Signing Gesture Markup Language," and XML-based version of the HamNoSys sign language writing system [Kennaway 2001]. This notational system was meant to be a writing/transcription system useful for researchers of signed languages, and it focuses on how to specify the handshape, palm orientation, and movement details significant for performing sign language. Compared to the motion control languages used in the previous two systems, the SGML is more tuned to the types of movements important for sign language performance. Therefore, the process of defining ASL signs using this notation should be more intuitive. To perform this animation, the designers of the ViSiCAST system use an animated human character that can accept SGML input and produce an animation. The ViSiCAST lexicon associated more linguistic information with each entry than merely its phonological SGML specification; special subcategorization, syntactic, and morphological features were also stored with each lexical entry.

The Workbench system uses a lexicon definition strategy that is even more sign language specific. This project has adapted the "Movement-Hold" model of American Sign Language phonology to create an electronic ASL sign representation format. The specification of ASL in the Workbench lexicon looks very similar to the Movement-Hold style of analysis currently prevalent in the ASL linguistic literature coming out of Gallaudet University. In this regard, the phonological specification of this system is the most expressive and well-suited to the needs of an ASL generator of any of the four surveyed systems. However, due to the highly detailed nature of the Movement-Hold model, the process of defining lexical items in this system would be extremely work-intensive. Also, this system never implemented an animation component to perform the

sign specification – an important next step in this system's development and a true measure of its phonological success.

Degree of User Intervention Needed for Translation

While two of the systems in the survey, namely TEAM and ZARDOZ, are traditional machine translation systems that attempt to automatically convert English text into ASL animation without the help of a human translator, the other two systems are actually designed to take direction from a human operator. The ASL Workbench system was designed to be a tool to assist professional translators convert from English to American Sign Language. In fact, the system is incapable of operating correctly without human intervention. It makes no attempt to resolve discourse referents in an English text; therefore, it requires the user to note whenever two references point to the same discourse entity. While ViSiCAST is less extreme in its need for human intervention, the designers of the system did give users the option of intervening between the pipeline stages of the translation process to correct any errors the system had made up to that point of processing to prevent their further propagation. Unfortunately, the publications of this project do not describe how often this intervention was necessary in order for the system to produce correct ASL output.

The need for a human operator to intervene in the translation process limits the applications for which these systems can be employed. Human-assisted machine translators could be useful for preparing ASL animations to be stored and transmitted over webpages or over television closed captioning systems. Such translators could also be useful for deaf individuals who were using a translation device to "word process" ASL animations. In both of these cases, the operator of the product would need to be fluent in both English and ASL. For this reason, human assisted ASL translators would be inappropriate for educational applications for deaf students with early English skills or hearing users wishing to learn ASL. Real-time translation systems for English-to-ASL chat relay or handheld voice-to-text-to-ASL devices would also not be possible.

* * *

Greatest Strengths of the Four Systems

The danger in using point-by-point comparisons is that they can fail to address the particular ways in which each of the projects under consideration has distinguished itself from the rest. This section of the report will therefore discuss each system in terms of a particular strength or successful feature that is part of its design. For the proposed ZARDOZ architecture, this strength is its understanding of the need for spatial reasoning in the production of classifier predicates. Workbench distinguishes itself in its use of a highly expressive phonological representation, and similarly ViSiCAST in its use of a detailed Discourse Representation Structure. The unique emotive and expressive capabilities of the Human Modeling and Simulation lab's virtual human characters has enabled the TEAM project to excel in the production of adverbial expression and morphological operators.

ZARDOZ and the Use of Reasoning During Translation

An important contribution of the ZARDOZ architectural proposal was that the system designers insisted that successful translation from written to signed languages will eventually require sophisticated spatial commonsense reasoning in order to understand an English input and present it via the spatial metaphors and three-dimensional classifier predicated expressions of American Sign Language. They discuss how traditional MT translation approaches which side-step semantic understanding will only be able to produce ASL constructs up to a point. Eventually in order for a translation system to use classifier predicates and certain locative verbs correctly, the system will need to manage and arrange elements in a visual scene.

As a partial workaround to the need for such sophisticated AI reasoning systems, the ZARDOZ schemata-filling architecture could be used to produce these visual/spatial constructs. For example, the "next to" relation in an English sentence could default to a schema for the signer positioning the first object immediately to the right of the second in front of the signer at chest-level; "next to" in other contexts could trigger other appropriate schemata. Obviously, such an approach is still quite labor intensive and appropriate only for limited set of contexts.

An intriguing and ambitious proposal made by the ZARDOZ authors is that a future ASL generation system could use metaphorical reasoning or sign/world ontologies to creatively invent new signs when lexical holes are encountered during a generation process. For example, the authors discuss how fingerspelled A could be combined with the sign MEDICINE to produce a sign for "aspirin" at runtime. In this case, an ontology could tell the system that medicine is a supertype of aspirin, and it could use a letter-initializing algorithm to invent the sign. Metaphorical reasoning about spatial properties such as “up is to down as improve is to worsen” could also be used to invent new signs; for instance, the sign for money in a downward motion could be used to derive a sign for "recession." Obviously such reasoning would be complex and involve a strong understanding of the semantics of particular English lexical items. These techniques are currently beyond the reach of technology, but they represent an interesting glimpse into the future directions ASL generation might take.

Workbench and the Model of ASL Phonology

An attractive feature of the Workbench system is its detailed and linguistically-motivated phonological representation which employs the Movement-Hold model of ASL signs [Liddell & Johnson 1989]. The system's output is therefore a series of "segments," atomic time-slices of ASL performance (smaller than a single sign) in which the details of handshape, location, orientation, non-manual signals, timing, movement, and other information are fully specified. Since much of this information is determined by grammatical constraints, feature expression, and morphological and phonological transformations on signs, the ASL-Generation lexicon entry for a sign may be underspecified, leaving holes in the segments that must be filled during generation. So, the entry for a sign is not considered a final piece of output but rather just a starting point for manipulations on the segment stream during generation.

The advantage of this highly detailed phonological representation is that it is able to capture all of the manual and non-manual forms of expression currently understood by modern ASL linguists. It also makes it easier to specify morphological and phonological rules, many of which seem to operate naturally on a Movement-Hold segments. In fact,

Speers formalizes many of the notational transformations his representation undergoes during several kinds of morphological and phonological operations.

The downside of this phonological formalism is also related to its level of detail; it's very difficult and sometimes cumbersome to fully specify all of the performance features of output for every time-slice. Building such a detailed lexicon would be resource-intensive. Another issue is that the Workbench lexicon entries hold little more information than this phonological root; grammatical feature bits representing agreement, morphological properties, and other information that would useful when implementing a robust grammar seem to be absent. While the use of Movement-Hold style phonology gives the lexical approach of the Workbench system a definite advantage, there are still many ways the system would need to be enhanced in order to produce a successful ASL generation lexicon.

ViSiCAST and the Discourse Representation Structures

The Discourse Representation Structure (DRS) formalism [Sáfár and Marshall 2002] chosen for the ViSiCAST system is a good choice as the semantic bridge between the English analysis side and the ASL generation side of the transfer-based system. The DRS explicitly lists all of the referents in the discourse, which is important since management and tracking of discourse entities is critical to ASL's pronoun and verbal agreement expression. The authors correctly asserted that ASL's particularly unambiguous pronouns (gesturing to the point in space associated with a particular entity) require the translation process to accurately determine coreference in the source language; so, the authors also began work on such reference resolution algorithms for their system. While the DRS's explicit tracking of discourse entities is useful for an English-to-ASL system, noticeably lacking from the DRS for ViSiCAST was information about the specific point in space which has been selected for each entity. This important feature was not implemented in the initial version of the system.

In the semantic portion of the DRS, information is stored as sets of simple propositions. This aspect of the formalism is particularly useful for an English-to-ASL system because it helps to break apart highly aggregated English sentences into sets of individual predicates – a more natural starting point for a language with typically short

sentences as in ASL. This decomposition also helps to reduce any English-syntax-bias of the semantics stored in the DRS; the isolation of tense, aspect, and other modifying phenomena in the DRS is important since ASL may express these using different grammatical constructs or modalities [Sáfár and Marshall 2002].

Another way in which the DRS semantic representation is well-suited to ASL is that it includes higher-order predicates which can act on other labels in the representation. The authors point out that such predicates are useful for representing information such as negation, adverbials, or verb modification that is often communicated via the NMS channels parallel to the manual sign stream.

TEAM and the EMOTE Motion Parameterization

A complex problem for a generation or machine translation system is that much of the aspectual and adverbial information in ASL is communicated via subtle morphological variations on individual signs or changes in the non-manual expressions which accompany the signing performance. Little aspectual information and few of the most common adverbial modifications are lexicalized in ASL. Deciding how to convert lexical adverbial information from an English source text into facial expressions or minor variations in the performance of a sign can be difficult without populating the lexicon with dozens of alternate performances for each possible sign (under each possible adverbial/aspectual environment) or cluttering the grammar with directives to control detailed aspects of the non-manual performance (the grammar writer should not have to write rules directing how high to raise each eye-brow).

To keep these complexities under control and to make the problem of expressing adverbial/aspectual information in an ASL generator a reasonable one to tackle, an initial model of how ASL conveys this information was developed for the TEAM system. This system used the EMOTE motion parameterization component [Badler et al. 2002] of the Jack Visualization system to generate some ASL adverbials. The adverbial expressions of particular interest were those involving temporal aspect, manner, and degree. The EMOTE framework was also successfully used to implement several types of inflectional and derivational morphological operators requiring similar nuanced modifications to the performance of a root sign morpheme.

The Jack animated virtual human is able to plan and execute a set of movements using a planning architecture that manipulates operators called Parameterized Action Representations [Badler et al. 2000]. To enable the virtual human character to select the manner by which it will perform its motions, the designers created a movement manner parameterization scheme. The Laban theory of human motion characteristics was used a basis for the design of this framework; in this theory, all human motion can be described in terms of values on continua grouped into five categories: Body, Space, Effort, Shape, and Relationship [Badler et al. 2002]. The Effort and Shape categories are of particular interest in conveying what we perceive as the "manner" of a human's motion, and so the EMOTE designers chose to implement the Effort and Shape parameterizations of human motion from the Laban theory in their system.

As the motion planning system produces a movement script for the animated character to follow, it can augment the motion primitives with EMOTE parameterization values to affect the manner of performance. These few parameters can be used to control and coordinate multiple complex channels of gestural and facial expression. The designers of the system have produced an adverb lexicon that stores a set of EMOTE parameters with each entry that produce an animation which typifies that adverb. Thus, by adding a single “adverb” to the movement script, and entire series of animation nuances are affected which produce a movement in the style of that adverb. While this EMOTE adverb lexicon was initially produced to support the natural language command and control of a virtual human character, the designers of the TEAM system adapted this technology so that the entries in the adverb list could be used to modify sign language performance in a manner which would express the meaning of an adverb via non-manual channels.

Since it was based on STAG, the heart of the TEAM translation system was a bilingual lexicon in which each of the two corresponding English/ASL entries was associated with a TAG representing how the lexical item interacted with the syntax of its language. Each ASL sign in the lexicon pair also stored a set of virtual-human-compatible action primitives which instructed the system how to animate a character to perform the sign; this representation was augmented with a set of EMOTE parameters which were appropriate for the standard way in which the sign was performed. As an

adverbial modifier was added to the growing ASL linguistic tree structure during the English-to-ASL transfer process, it could take on several forms. Some adverbials were associated merely with ASL signs (and the appropriate EMOTE parameters for that sign alone), others were associated with a sign and a set of EMOTE features that spread through the tree, and others had the spreading features but no lexical sign. By the term "spreading," it is meant that the process of adding the adverbial to the tree structure initiated an operation whereby the EMOTE parameters for other signs in the tree were modified so that the resulting animation would convey the semantics of the particular adverbial.

This TEAM approach to adverbials is attractive because it parameterizes the control of the animated character in order to simplify the task of defining a new adverbial in the lexicon. The fact that the parameterization scheme is based on the respected movement description standard of Laban Analysis also gives the approach additional credibility. Using this EMOTE-based technique, TEAM is the only surveyed system to handle complex non-lexicalized ASL adverbial expression.

* * *

Conclusion

In many ways, this report has likened the English-to-ASL problem to traditional machine translation language pairings, and indeed, all of the systems in the survey incorporate some standard computational linguistic tools and formalisms. The TEAM system uses the Synchronous TAG translation approach with the XTAG grammar for English analysis, and the ASL Workbench uses the Lexico-Functional Grammar formalism. The ViSiCAST translator makes use of the CMU Link Parser, Discourse Representation Theory, and a Head-Driven Phrase Structure Grammar. Even the rather unorthodox ZARDOZ system makes use of the Doll Control Language and Spatial Dependency graph grammar – both of which were developed prior to that system.

The more interesting findings of this survey surround those design choices made in the four systems because the special properties of American Sign Language have required the development of new computational linguistic technologies. New sign language symbolic representation systems have been developed for several projects: TEAM uses

an extended "sign gloss" formalism, ViSiCAST compiles a similar formalism into HamNoSys-style output, and the ASL Workbench computerizes the Movement-Hold phonological model. This report has discussed how the symbolic representation chosen for the ASL output can subtly limit the expressiveness of the eventual output – especially for the important NMS channel. The TEAM system has also demonstrated how the ASL representation should make it easy to parameterize modifications to the standard sign performance to enable the expression of adverbial information or other forms morphological operators in ASL.

Systems like ViSiCAST and Workbench have explored how the discourse representation should be adapted to manage the spatially positioned entities in an ASL dialogue. However, only the ZARDOZ system has begun to address how the spatially complex and three-dimensionally representative system of ASL classifier predicates could be produced from an English input text. That project has shown that a special form of scene representation and spatial processing is needed to handle classifier predicates, locative or directional verbs, and other complex uses of the signing space.

This report has also discussed how the grammar formalism and generation algorithms chosen for ASL should take into account the special properties of the language. An ideal approach should allow simultaneous access and modification to multiple channels of output, and it should take advantage of the word order flexibilities of ASL to instead focus on the other many complex choices to be made during generation. The discussion of system architectures has also illustrated how the divergence handling power vs. development effort trade-off of the MT pyramid is particularly acute for ASL since the lack of corpora in this language prevents the gathering of statistical information for machine translation.

* * *

New Directions

As a minority language whose advances in linguistic understanding are quite recent, American Sign Language is new to the field of machine translation, and so efforts to develop ASL MT technology are still continuing. Of the few systems that have attempted to develop natural language generation or translation technologies for ASL, only the

ViSiCAST system in this survey continues as a sustained research project. In addition to that project's efforts to extend its syntactic coverage, there are several active projects focused on developing computational linguistic components that would be useful to ASL translation system designers. There are efforts at producing ASL sign lexicons [Furst et al. 2000], collecting sign information via motion capture technology [Bangham et al. 2000], developing more sophisticated models of the human form for sign performance [Davidson et al. 2001], and collecting annotated corpora of ASL [Neidle 2000]. There are now several commercial systems [iCommunicator 2003] [Vcom3D 2000] that can produce Signed Exact English output from an English source text, and the Vcom3D Sign Smith Studio product allows users to edit the animation by hand to produce signing which is more ASL-like in structure. Finally, there is an English-to-ASL machine translation project growing out of the work of the TEAM system and the Human Modeling and Simulation technology here at the University of Pennsylvania [Huenerfauth 2003]. While still a newcomer to machine translation, ASL is a language with a bright future in computational linguistic research.

References

N. Badler, R. Bindiganavale, J. Bourne, M. Palmer, J. Shi, and W. Schuler. 1998. "A parameterized action representation for virtual human agents." In Workshop on Embodied Conversational Characters, Lake Tahoe, CA. http://www.cis.upenn.edu/~rama/publications.html

N. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao, S. Lee, H. Shin, and M. Palmer. 2000. Parameterized Action Representation and Natural Language Instructions for Dynamic Behavior Modification of Embodied Agents. AAAI Spring Symposium. ftp://ftp.cis.upenn.edu/pub/graphics/rama/papers/aaai.pdf

N. Badler, J. Allbeck, L. Zhao, and M. Byun. 2002. "Representing and Parameterizing Agent Behaviors." http://www.cis.upenn.edu/~badler/papers/ca02.pdf

J. A. Bangham, S. J. Cox, R. Elliot, J. R. W. Glauert, I. Marshall, S. Rankov, and M. Wells. 2000. "Virtual signing: Capture, animation, storage and transmission - An overview of the ViSiCAST project." IEEE Seminar on "Speech and language processing for disabled and elderly people."

M. J. Davidson et al. 2001. Improved Hand Animation for American Sign Language. Technology And Persons With Disabilities Conference.

B. Dorr, P. Jordan, and J. Benoit. 1998. "A Survey of Current Paradigms in Machine Translation." http://citeseer.nj.nec.com/555445.html

J. Furst, K. Alkoby, A. Berthiaume, P. Chomwong, M. J. Davidson, B. Konie, G. Lancaster, S. Lytinen, J. McDonald, L. Roychoudhuri, J. Toro, N. Tomuro, R. Wolfe. 2000. "Database Design for American Sign Language." Proceedings of the ISCA 15th International Conference on Computers and Their Applications (CATA-2000). 427-430.

J. Holt. 1991. Demographic, Stanford Achievement Test - 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results.

M. Huenerfauth. 2003. "Incorporating Research from Virtual Human Modeling into the Study of American Sign Language Generation and Machine Translation." Research Report, Independent Study, Computer and Information Science, University of Pennsylvania.

iCommunicator 4.0 Website. 2003. http://www.myicommunicator.com/

J. R. Kennaway. 2001. "Synthetic animation of deaf signing gestures." 4th International Workshop on Gesture and Sign Language Based Human-Computer Interaction, London. Lecture Notes in Artificial Intelligence vol. 2298 (eds. Ipke Wachsmuth and Timo Sowa).

S. Liddell and R. Johnson. "American Sign Language: The Phonological Base," Sign Language Studies, 64, pages 195-277, 1989. In C. Valli & C. Lucas, 2000, Linguistics of American Sign Language, 3rd edition, Washington, DC: Gallaudet University Press.

I. Marshall and É. Sáfár. 2001. "Extraction of semantic representations from syntactic SMU link grammar linkages." In G. Angelova, editor, Proceedings of Recent Advances in Natural Lanugage Processing, pages 154-159, Tzigov Chark, Bulgaria, September.

C. Neidle, J. Kegl, D. MacLaughlin, B. Bahan, and R. G. Lee. 2000. The Syntax of American Sign Language: Functional Categories and Hierarchical Structure. Cambridge, MA: The MIT Press.

C. Neidle. 2000. SignStream™: A Database Tool for Research on Visual-Gestural Language. American Sign Language Linguistic Research Project, Report Number 10, Boston University, Boston, MA, 2000.

S. Prillwitz, R. Leven, H. Zienert, T. Hanke, J. Henning, et al. 1989. "Hamburg Notation System for Sign Languages - An Introductory Guide." International Studies on Sign Language and the Communication of the Deaf, Volume 5, Institute of German Sign Language and Communication of the Deaf, University of Hamburg.

É. Sáfár and I. Marshall. 2001. "The architecture of an English-text-to-Sign-Languages translation system." In G. Angelova, editor, Recent Advances in Natural Language Processing (RANLP), pages 223-228. Tzigov Chark, Bulgaria.

É. Sáfár and I. Marshall. 2002. "Sign language translation via DRT and HPSG." In A. Gelbukh (Ed.) Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics, CICLing, Mexico, Lecture Notes in Computer Science 2276, pages 58-68, Springer Verlag, Mexico.

d'A.L. Speers. 2001. Representation of American Sign Language for Machine Translation. PhD Dissertation, Department of Linguistics, Georgetown University.

C. Valli & C. Lucas. 2000. Linguistics of American Sign Language, 3rd edition, Washington, DC: Gallaudet University Press.

VCom3D. SigningAvatar Frequently Asked Questions.(2000) http://www.signingavatar.com/faq/faq.html

T. Veale, A. Conway, B. Collins. 1998. "The challenges of cross-modal translation: English to sign language translation in the ZARDOZ system" in Machine Translation 13. 81-106.

C. J. Wideman and E. M. Sims. 1998. "Signing Avatars." Technology And Persons With Disabilities Conference.

L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer. 2000. "A Machine Translation System from English to American Sign Language." Association for Machine Translation in the Americas.