The Architecture of an English-Text-to-Sign-Languages Translation System

Eva Safar and **Ian Marshall**
School of Information Systems
University of East Anglia
Norwich NR4 7TJ, Great Britain
{es,im}@sys.uea.ac.uk

Abstract

We present an overview of the overall architecture of the language-processing component of an English-Text to Sign-Languages translation system[1]. Initially relevant aspects of sign languages and the influence these have on the design decisions in the translation system are described. Focusing on the analysis stage, we then present the syntactic, semantic and discourse oriented natural language processing (NLP) techniques, that are implemented to generate the DRS-based semantic representation from English text. We also give an account on the pronoun resolution algorithm, which is tightly coupled with the DRS creation.

1 Introduction

ViSiCAST is a 3-year project funded as part of the EUs Framework V programme. The project develops virtual signing technology in order to provide information access and services to Deaf people. ViSiCAST's concerns include investigation of sign language delivery using different technologies (Elliott *et al.* 00), investigation of the potential of speech to sign language translation in restricted domains (Cox *et al.* 01), investigation of a multilingual sign translation system designed to translate English text into several European sign language variants.

The system described in this paper is a multilingual sign translation system designed to translate from English text into a variety of national sign languages (NGT (Dutch), DGS (German) and BSL (British)) in contrast to other text-to-sign-language translations systems like VCom3D (Wideman 00), Simon (Elliott *et al.* 00), which have been implemented to present textual information as SE (Signed English) or SSE (Sign Supported English): SE uses signs in English word order and follows English grammar, while SSE signs only key words of a sentence. The Tessa system (Cox *et al.* 01) translates from speech to BSL, but is built on a quite inflexible template-based grammar. Furthermore, an advantage of DRS-based semantic approach of the current system over the interlingua approach of the Zardoz system (Veale *et al.* 98) is a higher modularity by which a language independent grammar development for the target language is allowed. Additionally it is not only attractive for restricted domains, to which interlingua systems are most well suited.

English text to sign language translation is decomposed into two major stages, manipulation of the English text into a semantic-based representation and secondly translation from this representation to graphically oriented representations which can drive a virtual avatar.

The latter strand of this research is concerned with comparison and evaluation of different data sources as the basis for the sign language synthesis stage. Sign translation raises a number of alternatives for the synthesis stages ranging from smoothed concatenation of motion captured data from a sign dictionary (Elliott *et al.* 00) through to synthesized hand, face and body motion derived from the parallel and sequential composition of morphological sign primitives (each of which may be motion captured or may be synthesized using a sign gesture based notation) (Kennaway 01).

Research at the IDGS (University of Hamburg, Germany), IvD (Netherlands) and UEA (Norwich) is concerned with supporting semi-automatic preparation of English text for signed presentation in German, Dutch and British sign language respectively. IDGS and UEA are concerned with refinement of a Sign Language Notation and visualization in a virtual avatar (Kennaway 01). Each establishment develops lexicons and grammar synthesis rules for the respective national sign language (at UEA this is in conjunction with the UK RNID). Translation of English text to an intermediate semantic representation is researched at UEA and is the focus of the current discussion.

This paper reports on design decisions justifying the semantic representation, the employed NLP techniques and the interface from English text to sign language synthesis stages. Section 2 briefly describes relevant aspects of sign languages, which challenge a translation system. Section 3 is devoted to the overall text processing architecture. Sections 4, 5 and 6 describe the syntactic parsing, translation to the semantic representation, and the pronoun resolution stage respectively. Current progress in the realization of the natural language component is also outlined in Section 7.

[1]This work is incorporated within ViSiCAST, an EU Framework V supported project which builds on work supported by the UK Independent Television Commission and Post Office.

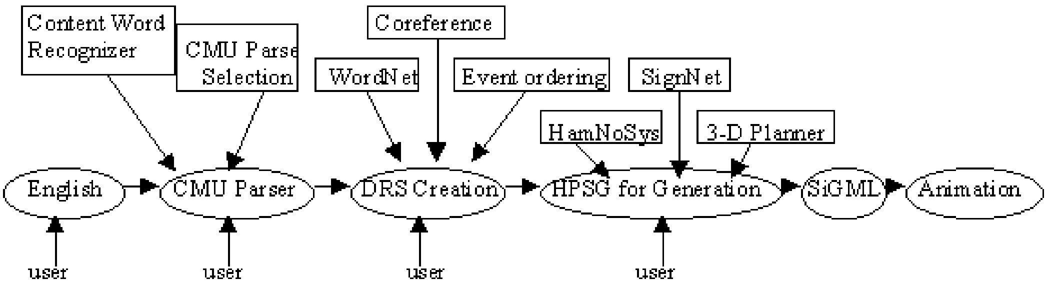

Figure 1: Stages of English text translation to sign language

2 Sign Language Features

Natural sign languages have a number of similarities to oral natural languages, though the three dimensional nature of the space around a signer afford a number of opportunities unavailable to oral languages. The following component descriptions of major features of British Sign Language (BSL) are based on (Brien 92) and (Sutton-Spence & Woll 99). In Section 4 the choice of the semantic representation is explained that reflects these main characteristics of BSL (British Sign Language).

- Sign Order

BSL has a topic-comment structure, in which the main informational subject or topic is signed first. The flexibility of sign order is due to the extra information carried in the directional verbs (see later) and eye-gaze.

- Signing Space, Placement and Pronouns

In a discourse, components of a description can be situated in the space in front of the body of a signer: first the area is defined and then all items or actions are related to that area. This means also that English is underspecified when using plural pronouns, while BSL can express the following: WE-TWO, WE-THREE, etc and distinguish between inclusion or exclusion of the 'hearer' (the communicating non-signer).

- Directional or Agreement Verbs

Agreement verbs include the information about person and number of the subject and object. This is realized by moving the verb in the syntactic space. The signing of the verb begins at the position of the subject and ends at the position of the object(s) (GIVE, TELL, etc), some verbs begin at the object and finish at the subject (BORROW).

- Classifiers

Classifiers are handshapes that can denote an object from a group of semantically related objects. The handshape is used to denote a referent from a class of objects that have similar features. (BICYCLE-PASS).

- Time lines

BSL has no tense system. Rather than express temporal information by morphological or syntactic features associated with verbs, it is expressed with the help of time lines in the signing space, by the ordering of the propositions or temporal adverbials.

3 Text Processing Architecture

The organization of the English text language processing component of the current system is shown in Figure 1. This is organized as a collection of automatic transformation components augmented by user-interaction. Human intervention is required to enhance the quality of the translation when automatic techniques are insufficient.

In the first stage, the user is allowed to change original text to rephrase unsupported constructions prior to processing. In the syntactic stage, the text is parsed by the CMU (Carnegie Mellon University) link grammar parser (Sleator & Temperley 91). During this stage the user can manually intervene to correct part of speech assignments and select between possible parse analyses.

From the selected link grammar parse, an intermediate semantic representation is built in the form of a Discourse Representation Structure (DRS). In addition, manipulation of the semantic representation allows for word sense assignment (via WordNet (Miller *et al.* 93) and/or manual intervention), determination of co-reference relationships and semantic reorganization to conform to linear time ordering of events (see section 2 about time lines).

The morphology and syntax of sign-generation from this semantic representation is defined within the framework of Head-Driven Phrase Structure Grammar (HPSG). At this stage signs can be edited by changing morphemes (e.g.: movement or handshapes) in the sign language grammar. This linguistic analysis is then linked with the animation technology via a Signing Gesture Markup Lanugage (SiGML), that is an XML-compliant representation of gestures (Elliott *et al.* 00) and is based on the refined HamNoSys (Prillwitz *et al.* 89) sign notation.

4 Parsing

The English text is the input for the CMU parser (Sleator & Temperley 91). Part of this grammar is a dictionary which defines the links with which a word must make to other words to be incorporated within a sentence. The parsers output is a set of links - a linkage - for a sentence. The CMU parser as a shallow analyzer in MT contexts has been chosen, as it is robust

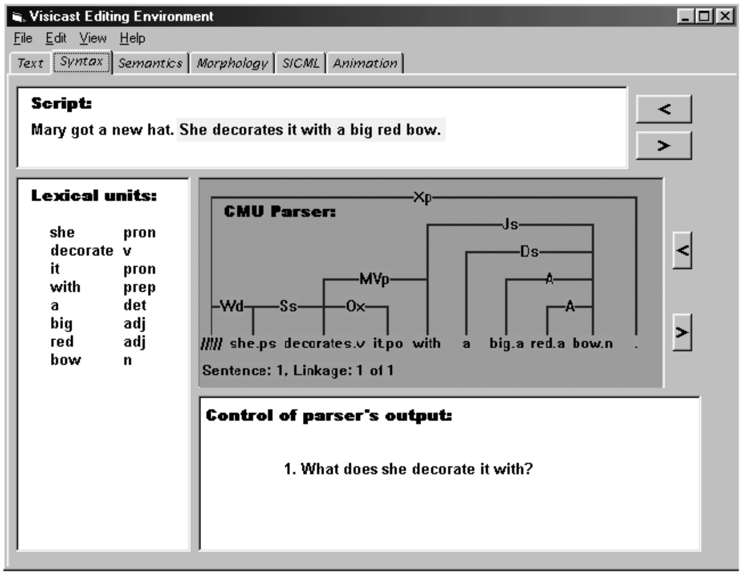

Figure 2: The script and the output of the CMU parser in an envisaged editing environment

and covers a significantly high proportion of English linguistic phenomena. For further details how the flat structure of the CMU output is used for building the semantic representation, see Section 5.

The CMU parser's dictionary has been modified to improve the handling of relative clauses with proper nouns, some phenomena of the punctuation, to permit part of speech assignment to verbs (such as 'be' and 'have' as auxiliaries) and to handle contracted forms ('hasn't') and non-contracted forms ('has not') consistently.

The parser often produces a high number of linkages for one sentence. Currently the user has to select the correct linkage by direct intervention. However, it is foreseen that selection between alternate linkages can be achieved by questions which guide linguistically less sophisticated users to the correct analysis.

Figure 2 illustrates the envisaged editing environment of the system with the currently available CMU parser output and script slots.

5 Semantic Representation

The approach to English to Sign Language translation is based upon use of Discourse Representation Theory (DRT) (Kamp & Reyle 93) for the intermediate representation of meaning. A DRS (Discourse Representation Structure) is a two part construction involving a list of variables denoting the nominal discourse referents and conditions (a collection of propositions which capture the semantics of the discourse). DRT was chosen as the underlying theory because it decomposes linguistic phenomena into atomic meaning components (propositions with arguments), and hence allows isolation of tense/aspect and anaphoric connections that are realized in different sign language grammatical constructs or modalities (see Section 2).

DRSs, described in (Kamp & Reyle 93), are modified to achieve a more sign language oriented representation that subsequently supports an easier mapping into a sign language grammar. In (Kamp & Reyle 93) only event propositions are labeled for use as arguments with temporal predicates. This has been extended by introducing labels for each semantic predicate. As in Verbmobil the labeling of all semantic entities allows a flat representation of the hierarchical structure of arguments and operator embeddings(VIT representation) (Dorna & Emele 96; Dorna 00). The label arguments referring to other elements in the flat list are useful to handle verb modifiers, negation and adverbials (e.g.: [attr1:big(X), attr2:very(attr1)]), which reflect the multiple modalities in sign languages: facial expressions convey intensity, head nod (parallel to signing) negation. This form of representation also has the advantage, as Dorna & Emele claim, that additional constraints which are important for generation in the target language, e.g. topic/focus in sign languages, may be made explicit. Additionally, in contrast to Vermobil's uniform labeling an ontology for all DRS propositons has been introduced to ease the mapping between the flat semantic structure of the DRS to the input structure of the target language specific HPSG, as required by the generation algorithm in ALE (Carpenter & Penn 99). The class definitions, that result from the taxonomy of labels, can help with the search of possible anaphoric referents and temporal relationships (See Section 2 for timelines).

The translation from a CMU linkage to its DRS representation happens via Definite Clause Grammar (DCG) rules, which is implemented in Prolog. The CMU parser's flat output structure, the links are ordered on a fixed preferential basis according to their start and end positions and irrelevant (redundant)

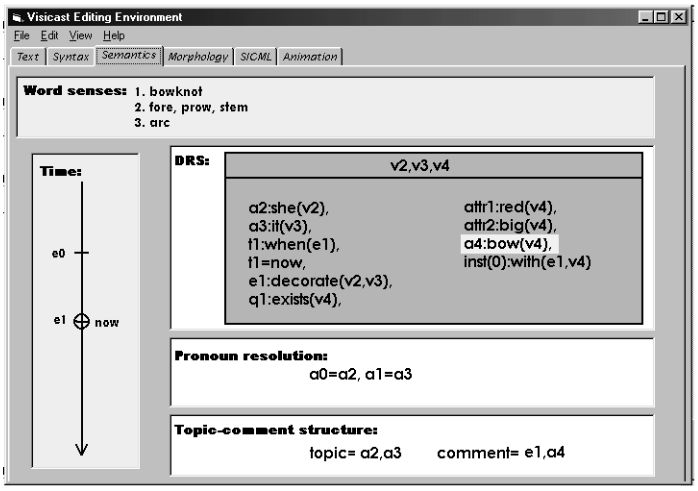

Figure 3: The output of the DRS for the second sentence and the result of the pronoun resolution in an envisaged editing environment

links are deleted. A link dictionary maps each link type to a λ-expression DRS definition (λ-DRS). The DCG then concatenates the λ-DRSs in the right order and steers the constants into the correct slots of the predicates, realizing the hierarchical order of arguments and semantic operators (Blackburn & Bos 99). After using functional application (β-reduction) the merge operation combines the DRSs by making the union of the universes and conditions as in (Bos *et al.* 94).

Figure 3 shows the Semantics window of an envisaged editing environment for the currently available DRS output.

6 Pronoun Resolution

In sign languages, pronouns are pointing gestures to the location associated with a noun (see Section 2). These languages make a very extensive use of this placement of referents at particular points in the signing space, therefore anaphora resolution in English text is crucial for a correct translation into BSL.

A significant number of antecedents which are potential referents for a pronoun in English can be excluded by linguistic restrictions on gender-number agreement, intra- and intersentential accessibility constraints in DRSs. These constraints are applied in a small window (2 sentences), and subsequently a robust approach is required for further resolution with the view to avoiding complex semantic and discourse analysis. This is vital for a translation system as a real-world application.

Antecedents that obey the above mentioned constraints are scored by preferences. The weighting algorithm is a modified version of the work by (Kennedy & Boguraev 96). They claim "the strong points of this algorithm is that it operates primarily on syntactic information alone" with 75% accuracy rate. The current implementation is an improvement to the gender agreement in (Kennedy & Boguraev 96) by augmenting the algorithm with a lexical database (female, male names, WordNet), to conditions on coreferents in (Kennedy & Boguraev 96) by making use of accessibility constraints in DRS. The modified algorithm improves on the suitability idea of (Kamp & Reyle 93) by determining how to choose from more than one referent in the DRS.

The original Kennedy/Boguraev algorithm calculated the salience weight of a possible referent as the sum of the salience factors (grammatical role, adjunct, embedding, current sentence and complement of a preposition, existential construction). As each CMU link has an entry in the link dictionary defining its associated lambda expression, rather than compute the salience value, it can be associated directly in its link dictionary entry. Following (Kennedy & Boguraev 96) a COREF class is the collection of linguistic elements that corefer in a text to describe the same discourse referent. A COREF salience is associated with each of these. When an anaphor has to be resolved, the COREF class with the highest salience is selected. This possible referent is then checked for agreement in number, gender and accessibility within the DRS. Number agreement is checked with a noun stemming algorithm (though this could be changed to get the number information from the CMU linkage, e.g.: *Dms*, where *s* means singular), the gender of nouns is looked up in a database with female and male proper names, and the possible gender of common names is searched for in WordNet. Potential referents that do not satisfy these requirements are removed. When a link between a discourse referent (which can be another anaphor) and the current anaphor is established, this becomes a member of that class and its salience value is set to the COREF value of the antecedent. For each new sentence the old COREF values of previous sentences are halved. This means that the salience of a COREF class increases in the textflow according to the frequency of subsequent anaphoric references to it and decreases otherwise.

Figure 3 shows an envisaged editing environment for the currently available DRS and pronoun resolution result of the following text:

```
Mary got a new hat.
She decorates it with a big red bow.
```

Though there are situations where the referent for a pronoun will be incorrectly selected this algorithm has the benefit that it incorporates many aspects of a natural interpretation of pronoun resolution based on linguistic structure as well as experimental results.

7 Current State and Future Work

Currently the translation system of English text into a DRS-based intermediate semantic representation handles the following linguistic phenomena: transitive, intransitive verbs, temporal auxiliaries, passive, imperative, infinite number of noun and verb modifiers, subject and object type relative clauses, prepositional phrases as adjunct of verb phrases and of noun phrases, determiners (numbers, demonstratives, universal, indefinite), polite requests, expletives, predicatives,pronouns, wh-questions, yes-no questions and negation. This is approximately a 50% coverage of the CMU grammar link, though these are involved in common syntactic constructions.

Pronouns can be resolved using algorithm described above. Since BSL makes an extensive use of placement in the 3-D space (see section 2), it is crucial for a correct translation that anaphora resolution is augmented by processing definite descriptions. This algorithm will be based on WordNet as the source for definite description resolution. Crucially, however this involves utilizing word sense disambiguation algorithms in order to resolve more profound forms of co-reference.

Currently it is also envisaged that each sentence will be annotated by a predicate 'comment(X)' indicating the topic-comment structure (see Sign Order in Section 2) to support the mapping to signing.

8 Conclusion

The modular architecture and utilization of existent linguistic knowledge resources such as the CMU link grammar parser and WordNet have facilitated design and implementation of the English text to sign language translation system. The isolation and identification of propositions relating to temporal phenomena, attribution, and the resolution of pronominal reference within the DRS representation contribute significantly to the further task of synthesizing this information into a sign language presentation.

The structure of the CMU linkage to DRS translation system is itself a modular architecture, consisting of the DCG based processing of the linkage, the dictionary of CMU link types to lambda expression definitions, and the embedding of lambda reduction and candidate anaphora weighting as semantic actions within the DCG. This has enabled an incremental development for a subset of English which can support parallel future work to synthesize sign language presentations for the currently supported subset, and extension of this subset to provide a more comprehensive system.

References

(Blackburn & Bos 99) P Blackburn and J Bos. *Representation and Inference for Natural Language. A First Course in Computational Semantics. Volume II.* http://www.coli.uni-sb.de/~bos/comsem/book1.html, 1999.

(Bos *et al.* 94) J Bos, E Mastenbroek, S McGlashan, S Millies, and M Pinkal. *A Compositional DRS-based Formalism for NLP Applications. Report 59.* Universitaet des Saarlandes, Saarbruecken, 1994.

(Brien 92) D (Ed.) Brien. *Dictionary of British Sign Language/English.* Faber and Faber, London,Boston, 1992.

(Cox *et al.* 01) SJ Cox, M Lincoln, MJ Nakisa, and J Coy. Tthe development and evaluation of a speech to sign translation system to assist transactions. In *Submitted to International Journal of Human Computer Studies, January 2001*, 2001.

(Dorna & Emele 96) M Dorna and M Emele. Semantic-based transfer. report 143. In *Proceedings of the 16th International Conference on Computational Linguistgics (COLING-96).* Copenhagen, 1996.

(Dorna 00) M Dorna. Semantikbasierter transfer - maschinelle bersetzung auf mengen linguistischer constraints. (doctoral thesis). In *Arbeitspapiere des Instituts fr maschinelle Sprachverarbeitung (AIMS). Vol.6, No.2*, 2000.

(Elliott *et al.* 00) R Elliott, JRW Glauert, JR Kennaway, and I Marshall. The development of language processing support for the visicast project. In *Assets 2000. 4th International ACM SIGCAPH Conference on Assistive Technologies.* New York, 2000.

(Kamp & Reyle 93) H Kamp and U Reyle. *From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Kluwer Academic Publishers, Dordrecht, 1993.

(Kennaway 01) JR Kennaway. Synthetic animation of deaf signing gestures. In *The Fourth International Workshop on Gesture and Sign Language Interaction, Gesture Workshop 2001 (GW2001)*, City University, London, UK, 2001.

(Kennedy & Boguraev 96) J Kennedy and E Boguraev. Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics COLING'96*, Copenhagen, Denmark, August 1996.

(Miller *et al.* 93) GA Miller, R Beckwith, Ch Fellbaum, D Gross, and K Miller. *An Introduction to WordNet. An On-line Lexical Database.* http://www.cogsci.princton.edu/~wn/, 1993.

(Prillwitz *et al.* 89) S Prillwitz, R Leven, H Zienert, T Hanke, J Henning, *et al. 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, 1989.

(Sleator & Temperley 91) D Sleator and D Temperley. *Parsing English with a Link Grammar.* Carnegie Mellon University Computer Science technical report CMU-CS-91-196, 1991.

(Sutton-Spence & Woll 99) R Sutton-Spence and B Woll. *The Linguistics of British Sign Language. An Introduction.* University Press, Cambridge, 1999.

(Veale *et al.* 98) T Veale, A Conway, and B Collins. The challanges of cross-model translation: English-to-sign-language translation in the zardoz system. *Machine Translation*, (13):81–106, 1998.

(Wideman 00) CJ Wideman. *Internet-enabled, 3D, SigningAvatar Software Offers Accessility for the Deaf.* http://www.signingavatar.com/nc/presentations/april_fose.html, 2000.