TALN 2001, Tours, 2-5 juillet 2001

Translation of English Text to a DRS-based, Sign Language Oriented Semantic Representation

Éva Sáfár (1) and Ian Marshall (2)

University of East Anglia
School of Information Systems
University of East Anglia
Norwich
NR4 7TJ
United Kingdom
(1) es@sys.uea.ac.uk
(2) im@sys.uea.ac.uk

Résumé - Abstract

Une vue d'ensemble de l'architecture de la partie linguistique d'un système de traduction de texte Anglais en langage de signes est présentée. Nous nous concentrons sur la traduction du texte de langue naturelle en une représentation sémantique pour l'aide a la traduction gestuelle. Nous rapportons nos progrès actuels dans l'application de techniques syntaxiques, sémantiques et contextuelles de traitement du langage naturel pour la génération de la représentation sémantique. Par ailleurs nous prenons en compte la necessité de fournir, a l'étape d'analyse du text, des moyens d'intervention manuelle pour améliorer la qualité de la traduction lorsque les techniques automatiques s'avèrent insuffisament précises.

An overview of the overall architecture of the language component of an English-Text-to-Sign-Languages translation system[1] is presented, focusing upon the translation of natural language text into a gesture-oriented interlingua representation. We report current progress in applying syntactic, semantic and discourse oriented natural language processing (NLP) techniques to generate the interlingua representation. In addition, an account is given of the provision for manual intervention in the text analysis stages to enhance quality of the translation when automatic techniques are insufficiently accurate.

[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. The project develops virtual signing technology in order to provide information access and services to Deaf people.

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, 2000), investigation of the potential of speech to sign language translation in restricted domains (Cox, 2000), investigation of a multilingual sign translation system designed to translate English text into several European sign language variants.

English text to sign language translation is decomposed into two major stages, manipulation of the English text into an interlingua representation and secondly translation from the interlingua 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, 2000) through to synthesised 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 synthesised using a sign gesture based notation) (Kennaway, 2001).

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 visualisation in a virtual avatar (Kennaway, 2001). Each establishment is concerned with development of 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 interlingua 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 interlingua semantic representation, and the pronoun resolution stage respectively. Current progress in the realisation of the natural language component is also outlined in Section 7.

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. However, the sign language of a geographical area has no direct relationship to its local oral language and may have features which are similar to those of geographically remote oral language. The following component descriptions of major features of British Sign Language (BSL) are based on (Magill, 2000) and (Sutton-

Spence, 1999). 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 topic is the framework within which the predication takes place. After the topic has been identified, the rest of the sentence is the comment, the new information on it. Furthermore BSL has no fixed order of basic elements(S,V,O). This flexibility is due to the extra information carried in the directional verbs (see later) and eye-gaze.

- Signing Space, Placement and Pronouns

In BSL (and many other sign languages) signers exploit the so called signing space in front of their body. In a discourse components of a description can be situated in that space: first the area is defined and then all items or actions are related to that area. Thus, BSL has more pronouns than English, which are articulated by pointing to a location previously associated with a noun. 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, in which the subject and the object are placed around the signer. The signing of the verb begins at the position of the subject and ends at the position of the object (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. They are used with verbs which require a classifier so that when combined with location, orientation, movement and non-manual features the composite forms a predicate. 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 four time lines in the signing space or by the ordering of the propositions in the discourse.

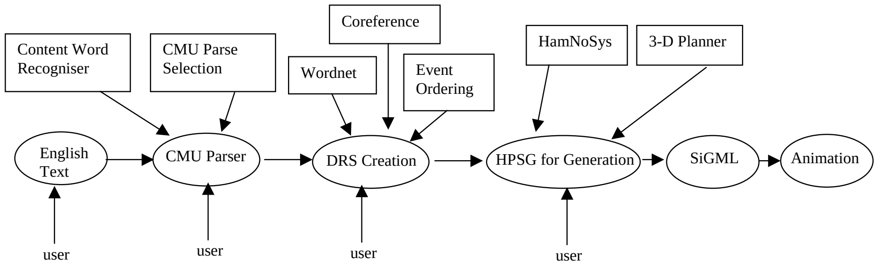

Figure 1: Stages of English text translation to sign language

3 Text Processing Architecture

The organisation of the English text language processing component is shown in Figure 1. This is organized as a collection of automatic transformation components augmented by user-interaction.

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, 1991). During this stage the user can intervene to correct part of speech assignments and select between possible parse analyses. From the selected link grammar parse, an intermediate interlingua 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, 1993) 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). Figure 2 illustrates the current state of the system.

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, 2000) and is based on the refined HamNoSys (Prillwitz, 1989) sign notation.

4 Parsing

The English text is the input for the CMU parser (Sleator, 1991). 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 is robust and covers a significantly high proportion of English linguistic phenomena.

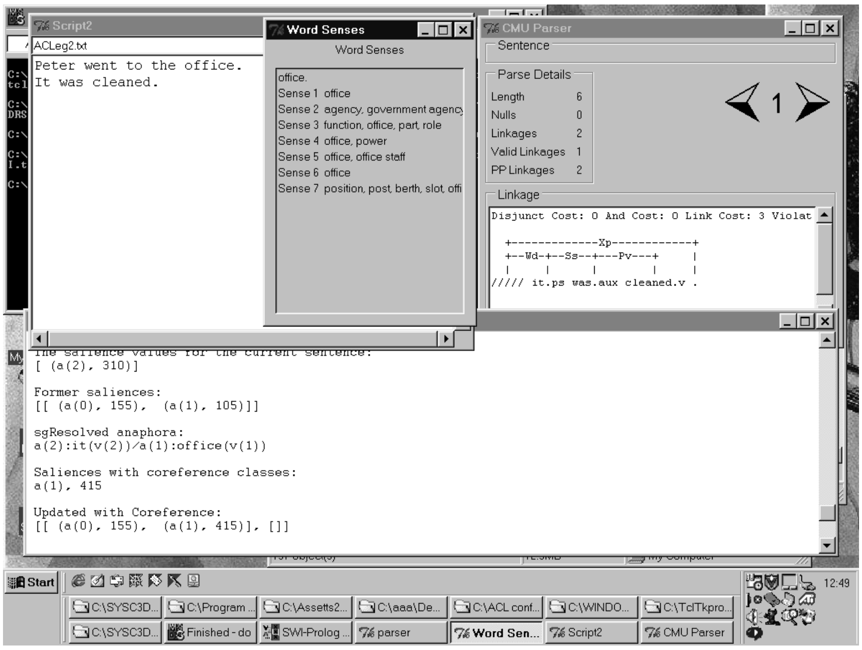

Figure 2: The current implementation

5 Semantic Representation

The approach to English to Sign Language translation is based upon use of Discourse Representation Theory (DRT)(Kamp, 1993) for the intermediate representation of meaning. 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 realised in different sign language grammatical constructs or modalities (see Section 2). DRSs, described in (Kamp, 1993), are modified to achieve a more sign language oriented representation that subsequently supports a more direct mapping into a sign language grammar.

A DRS 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). The translation from a CMU linkage to its DRS representation is implemented in Prolog.

In (Kamp, 1993) only event propositions are labeled for use as arguments with temporal predicates. This has been extended by introducing an ontology for all DRS propositons. This allows the possibility of [attr1:big(X), attr2:very(attr1)] to handle adverbial modifiers. In BSL facial expressions convey intensity and hence may be synthesised from such higher order predicates (for further modifications see also Section 7).

In the following example, the process of Discourse Representation Structure (DRS) creation is demonstrated. For Example 1 the CMU parser produces the linkage illustrated in Figure 3.

(1) Some nice men are walking in the park.

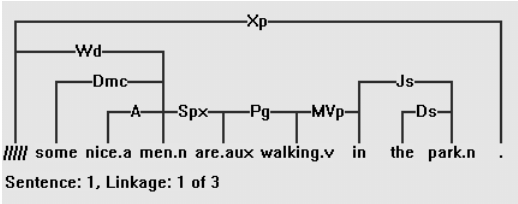

Figure 3: The output of the CMU parser

The linkage's links are ordered on a fixed preferential basis according to their start and end positions and irrelevant links are deleted. A link dictionary maps each link type to a λ-expression DRS definition (λ-DRS). These are λ-expressions, which may contain embedded DRSs or may take these as arguments. These λ-DRSs are concatenated to form a sequence of unevaluated λ-expressions, which is later reduced using functional application (β-reduction). At the end of the process, when no further β-conversions can be performed, a merge operation is carried out to build the complex DRS for the sentence as in (Blackburn, 1999). The merge operation combines two DRSs by taking the union of the two universes and the conditions (Bos, 1994).

Example 2 is the lambda expression associated with a 'Pg' link whose right end identifies verbs. Label indicates the kind of proposition, i.e. for event propositions e(Number1) and for temporal propositions t(Number2), where Number1 and Number2 are unique numbers. Verb stands for the verb-stem of the main verb identified by the link.

(2) Pg $\Longrightarrow$ lambda(X, drs([],[Label1:when(Label2),Tempus,Label2:Verb(X)]))

The determination of valid verb stems is achieved using WordNet (Miller, 1993) dictionary files. Similarly, morphologically available temporal information is extracted from auxilliaries and verbs (Tempus) and is identified by the temporal label (t) associated with the proposition as in Example 3.

(3) t(Number)=now, t(Number)=continuous

A context free Prolog Definite Clause Grammar (DCG) is used to process the linkage. Example 4 illustrates that the λ-DRS for np1 is the concatenation of λ-DRS for det applied to the λ-DRS for n2.

(4) np1(Sent, Det@Noun) $\longrightarrow$ det(Sent,Det), n2(Sent,Noun).

The link dictionary definitions for determiners and nouns provide λ-DRSs, which when reduced produce the λ-DRS in Example 5.

(5) drs([v(1)], [a(1):park(v(1))])

The link dictionary entries and the partially evaluated λ-DRSs can make explicit statements for merging, such as Example 6 for the phrase 'are walking in the park'.

```
(6) merge(drs([],
              [t(0):when(e(0)), t(0)=cont,t(0)=now,e(0):walk(v(0))]),
        drs([v(1)],
```

[a(1):park(v(1)), l(0):in(e(0), v(1))])))

The final DRS is shown in Example 7.

```
(7) sentence1
[v(0), v(1)]
[q(0):exists(v(0))
 attr(0):nice(v(0))
 a(0):men(v(0))
 t(0):when(e(0))
 t(0)=cont
 t(0)=now
 e(0):walk(v(0))
 a(1):park(v(1))
 l(0):in(e(0), v(1))
]
```

Though theoretically this is sufficient to implement DRS creation, link dictionary entries become rather complicated. Therefore, some DCG grammatical productions contain semantic actions to apply intermediate β-reduction to obtain subexpressions of the required form. This does not violate the compositionality principle and is beneficial for the current application in practical development of link definitions.

```
(8) v1(Sent,Label2,Beta) ⟶ auxp(Sent,(lambda(Aux,Tempus1,Aux))),
                           v(Sent,Label2,(lambda(X,drs([],
                                [Label1:when( Label2),Tempus2, Label2:Verb(X)])))),
    {
    Tempus1 =.. [Operator, Label1, T],
    betaConvert((lambda(Aux, Aux ) @
          lambda(X, drs([], [Label1:when(Label2),Tempus1, Tempus2,Label2:Verb(X)]))),Beta)
    }.
```

Example 8 illustrates the DCG production v1 for verbal phrases consisting of a sequence of auxillary verbs followed by a main verb and its embedded semantic action. The auxp and v DCG productions determine the λ-expressions associated with the appropriate links and the v1 production specifies that these are β-converted to generate the form of the λ-expression for v1 that is used in further β-reductions.

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.

```
sentence1

[v(0), v(1)]
[
  a(0):Peter(v(0))
  t(0):when(e(0))
  t(0)<now
  e(0):go(v(0))
  a(1):office(v(1))
  l(0):to(e(0), v(1))
]

The salience values for the current
sentence:
[ (a(0), 310),  (a(1), 210)]

Former saliences:
[]
Updated with Coreference:
[]
```

```
sentence2

[v(2)]
[
  a(2):it(v(2))
  t(1):when(e(1))
  t(1)<now
  e(1):clean(v(someone), v(2))
]

The salience values for the current sentence:
[ (a(2), 310)]

Former saliences:
[[ (a(0), 155),  (a(1), 105)]]

Resolved anaphora:
a(2):it(v(2))/a(1):office(v(1))

Saliences with coreference classes:
a(1), 415

Updated with Coreference:
[[ (a(0), 155),  (a(1), 415)], []]
```

Figure 4: The output of the DRS after pronoun resolution

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, 1996). 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, 1996) by augmenting the algorithm with a lexical database (female, male names, WordNet), to conditions on coreferents in (Kennedy, 1996) by making use of accessibility constraints in DRS. In addition the modified algorithm improves on the suitability idea of (Kamp, 1993) which does not determine how to choose the referent when more than one is available 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, 1996) 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 4 shows the DRS output with the pronoun resolution result of the following text:
(9) Peter went to the office. It was cleaned.

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 adjectives, subject and object type relative clauses, prepositional phrases as adjunct of verb phrases and of noun phrases, determiners and pronouns. This is approximately a 20% 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 utilising 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 utilisation 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, nominal 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 synthesise sign language presentations for the currenty supported subset, and extension of this subset to provide a more comprehensive system.

References

P. Blackburn, 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.

J. Bos, E. Mastenbroek, S. McGlashan, S. Millies, M. Pinkal. A Compositional DRS-based Formalism for NLP Applications. Report 59. Universitaet des Saarlandes, Saarbruecken, 1994.

SJ. Cox, M. Lincoln, MJ. Nakisa, J. Coy. The Development and Evaluation of a Speech to Sign Translation System to Assist Transactions. Submitted to International Journal of Human Computer Stuedies, January 2001.

R. Elliott, JRW. Glauert, JR. Kennaway, 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.

H. Kamp, U. Reyle. From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Dordrecht, Kluwer Academic Publishers, 1993.

JR. Kennaway. Synthetic Animation of Deaf Signing Gestures. The Fourth International Workshop on Gesture and Sign Language Interaction, Gesture Workshop 2001 (GW2001) at City University, London, UK. April 18th-20th, 2001.

J. Kennedy, 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, 5-9 August 1994.

R. Magill, A. Hodgson. Start to Sign!. The Royal National Institute for Deaf People, 2000.

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

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.

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

R. Sutton-Spence, B. Woll. The Linguistics of British Sign Language. An Introduction. Cambridge, University Press, 1999.