# Sign Language Generation using HPSG

Ian Marshall, Éva Sáfár School of Information Systems Norwich NR4 7TJ, United Kingdom im@sys.uea.ac.uk, es@sys.uea.ac.uk

#### Abstract

We discuss the problems of translating English to Sign Language in the ViSi-CAST<sup>1</sup> project. An overview of the language-processing component of an English-Text to Sign-Languages translation system is described focusing upon the inherent problems of knowledge elicitation of sign language grammar and its implementation within a HPSG framework.

## 1 Introduction

It is only since the mid 20th Century, sign languages have been recognised as 'natural' languages with their own phonology, morphology, syntax, semantics and pragmatics. Research over the past half century, however, has lead to greater recognition of sign languages and the Deaf communities in general (Klima 1979; Kyle & Woll 1985), and to an increased recognition as languages of interests to linguistic research. Arguably, it has also lead to a more sympathetic general public, as witnessed in the UK by the popularity of British Sign Language courses amongst the hearing population. Nonetheless, there remains a shortage of sign language interpreters within the UK at a time when legislation and directives require greater provision of information and service access for deaf people (Woll 2001).

'Virtual human' (avatar) technology has reached a stage where relatively realistic three dimensional human characters can be processed at sufficient frame (image) rates that signed presentations are readable by skilled signers. The performance increase and the cost reductions continue at a rate which suggests that such technology will not be prohibitively expensive in the future (Hennessy & Patterson 1996).

Our current research focuses on two main methods for signed content creation to be used to drive avatar technology (Elliott et al. 2000). One approach is based upon use of motion capture technology which permits the 'recorded' (motion captured) behaviour of a human signer to be stored and later replayed. This embodies realistic human motion by replaying signing derived from humans through an avatar (Cox et al. 2001) in a more flexible manner than simplistic videos. Significantly motion-captured signing retains synchronisation of dependent communication channels (facial expression, body posture, manual signing) of multi-modal communication.

 $<sup>^{1}</sup>ViSiCAST$  is an EU Framework V supported project which builds on work supported by the UK Independent Television Commission and Post Office. The project researches virtual signing technology in order to provide information access and services to Deaf people. The avatar shown in Figure 1 has been developed within the ViSiCAST project by our colleagues R.Elliott, J.R.Kennaway and K.J.Parsons.

However, motion-captured data is a relatively uninformed sign linguistics approach. The long term utility and more flexible approach is based upon a process of synthetic generation of signed presentation from linguistically motivated features (Kennaway 2001). This requires a knowledge-rich software environment for construction of a semantic representation from which signing can be generated. The text to signing research has initiated an approach to synthetic signing based upon a serious analysis and understanding of European sign languages (specifically German, Dutch and British sign languages).

# 2 English Translation and Sign Language Generation

During the past fifty years computational linguistic and natural language processing research has progressed significantly. However, even the most effective automatic MT systems (e.g. the web based Babelfish (Babelfish 2001)) produce target language output which is below the quality of human translation though indicative of the source language content (Hutchins 2001). Thus an environment is assumed which supports human intervention to volunteer information and correct automatic processing to enhance the quality of the final signed presentation. The system exploits and extends existing natural language processing technology for English in order to generate an appropriate semantic representation (Safar 2001; Marshall 2001).

The Carnegie Mellon University (CMU) parser (Sleator & Temperley 1991) syntactically processes English text to generate a parse linkage structure. The selected linkage is then tranformed into a Discoure Representation Structure (DRS) using Definite Clause Grammar (DCG) rules implemented in Prolog. A link dictionary maps each link type to a  $\lambda$ -expression DRS definition ( $\lambda$ -DRS) (Marshall 2001). The DCG then concatenates the  $\lambda$ -DRSs in the appropriate order and instantiates the arguments of the predicates appropriately (Blackburn & Bos 1999). Functional application ( $\beta$ reduction) and a DRS merge operation combine the  $\lambda$ -DRSs into the final DRS (Bos et al. 1994).

Discourse Representation Theory (DRT) (Kamp & Reyle 1993) uses a two part structure involving a list of variables (denoting the nominal discourse referents) and conditions (a collection of propositions which capture the semantics of the discourse). This decomposition of linguistic phenomena into atomic meaning components (propositions with arguments) in the condition allows isolation of tense/aspect and modifying phenomena that are realized in different sign language grammatical constructs or modalities. In addition, the centrality of co-referentiality in DRT addresses the need to appropriately determine how to assign fixed positions in signing space to significant discourse referents. The bottom middle window of Figure 1 illustrates the DRS structure for the sentence 'There are bowls in the sink.'.

The DRS representation has been modified for a more sign language oriented representation that subsequently supports an easier mapping into a HPSG sign language grammar. In (Kamp & Reyle 1993) only event propositions are labeled for use as arguments with temporal predicates, though singular and plural (v(N) and greek(N) in Figure 1) variables are distinguished. This has been extended by introducing labels for different kinds of semantic predicates (s-stative, e-event predicate, l-locative,



Figure 1: Input sentences, CMU Parser linkage, DRS and HamNoSys

a-nominative etc), and as with Verbmobil's VIT representation (Dorna & Emele 1996; Dorna 2000), the labeling of semantic entities allows a flat representation of the hierarchical structure of arguments and operator embeddings. The DRS representation is then converted to a nested HPSG semantic structure using the ontology of DRS propositions introduced in this labelling. In addition, this conversion handles transformation of differing numbers of complements between the English derived DRS events/states and the sign language oriented equivalents. In the example of Figure 1 the one argument predicate 'exist' and its locational adjunct are converted into a BSL two argument SEM structure containing the relation EXIST. The output of this stage is shown in Example (1) below.

# 3 Methodological Issues and HPSG development

Generation of a signed presentation from the semantic representation requires a sign language module which is informed about sign language syntax, morphology and phonology. The development of such a module is at the frontiers of sign language research.

Sign language research (as with other minority languages) poses a number of methodological issues. The ideal situation would be the development of a technological environment in which native signers can develop lexicons and grammars and explore the use of notations for describing their own language, and the longer term may hold this prospect. Fundamentally, this is a desire to reduce the 'design, implement, derive user feedback and revise cycle' to as short a time period as possible. However, the current use of notations (such as HPSG and HamNoSys) to describe formal properties of sign language is a highly specialist activity. In addition, the role of introspection has its critics, but in the absence of very large corpora in a suitable form to permit analysis of sign language in use, such insights are invaluable.

However, a fundamental question remains of whose insights are to be sought and what value is to be placed upon them. Sign research has frequently been carried out by hearing people using deaf informants and hence insights are typically second-hand. Additionally, the status of deaf informants themselves within the Deaf community raises a significant issue. Typically only 5-10% of deaf people are born to deaf parents and thus are viewed as the genuine native signers who should act as informants and who should be asked to identify the preferred manner of signing a proposition rather than merely acceptable signing (Neidle et al. 2000). Deaf informants with hearing researchers and initial review by hearing signers are used to establish initial hypotheses. More extensive review by deaf users of the generated signing provides detailed feedback and guides revision. Though far from ideal, this permits exploration of the use of the underlying formalisms prior to a more appropriate methodological framework.

#### 3.1 HPSG

The generation stage involves development of sign language grammars consistent with HPSG theory (Pollard & Sag 1994). This framework has not been used widely for generation, however a small number of projects have taken this approach (e.g. LinGO has been used to build a large-scale grammar for English using HPSG which is implemented in the Verbmobil machine translation system). Furthermore, the attempts to formulate sign language grammar have not typically elected to use the HPSG framework (Neidle et al. 2000), though some sign language constructs have been analysed in an HPSG framework (Cormier 1998).

However, variations in sign languages are less substantial in their grammars in comparison with their lexicons. Thus, a HPSG lexicalist approach is suitable for developing grammars for the three target languages in parallel. Differences are encoded in the lexicon, while grammar rules are usually shared with occasional variation in semantic principles. In addition, HPSG feature structures can incorporate modality-specific aspects (non-manual features) of signs appropriately.

A provisional BSL grammar has been developed using a Semantic Head-Driven (SHD) generator (Shieber et al. 1989) implemented in ALE (v3.2), an extension to Prolog<sup>2</sup>. It involves the standard components of an HPSG grammar, the feature structure specification, the lexicon, the grammar rules and principles. For each of these we characterise the significant aspects as they relate to the sign languages.

#### 3.2 Feature Structure

The feature structure is relatively large but consists of reasonably standard components like phonetic (PHON), syntactic (SYN) and semantic (SEM) structures. Much of the

<sup>&</sup>lt;sup>2</sup>Our German ViSiCAST partner explores the possibilities of LinGO (Copestake et al. 1999)

detail of the feature structure is focused on fine grain detail in the PHON component. An extension of the HamNoSys notation (Prillwitz et al. 1989) is used to describe the manual components of signs, handshape, palm orientation, finger direction and movement information. Generation is essentially oriented to determining these components for individual signs. Standard citation forms for signs are documented in various dictionaries, e.g. (Brien 1992), and provide appropriate information for initial descriptions of the manual components of signs. For fixed signs, whose phonological forms do not vary, this information is sufficient, but non-fixed signs require formulation of appropriate phonological and syntactic rules and principles in HPSG. As discussed above, elicitation of this knowledge poses difficult methodological problems.

The argument structure and the agreement components of the SYN structure determine conditions under which signs can be combined into a grammatically correct physical realisation. SEM structures include semantic roles and indexes as in LinGO. The latter proved to be necessary despite the nested goal definition in ALE to determine syntactic roles and agreement for a relatively free word order language (see also 3.3 and 3.4). The HPSG implementation is an initial formulation of such rules.

#### 3.3 The Generation Algorithm and the Semantic Input

ALE's internal generation algorithm is semantic head driven (SHD). It operates by discovering a pivot, which is the lowest node in a derivation tree that has the same semantics as the root.

Grammar rules are divided into two kinds, chain rules (which have a semantic head - whose head daughter's logical form is identical to the logical form of the mother) and non-chain rules (which have no semantic head or are lexical entries). The pivot is identified as the mother node of a non-chain rule operating in a top-down fashion. After the pivot has been found, it generates bottom-up using chain-rules to connect the semantic-heads to the pivot (Carpenter & Penn 1999).

Thus, this algorithm is requires a nested semantic (SEM feature) input structure illustrated in Example (1). In the remaining text we will use the term *semantic input* for such an input goal description:

Indices are introduced in the same way as with other generation algorithms such as the Shake-and-Bake algorithm (Copestake et al. 1995). Eventually these will be exploited for agreement and for associating discourse objects with particular positions in signing space for the purposes of co-reference.

### 3.4 Lexical Entries and Rules

ALE provides not only the type hierarchy declaration, format for lexical entries and the mechanism for unification, but also a way to change morphological realization of lexical entries using lexical rules. The standard ALE implementation generates a result which is a sequence of lexical items derived from the left hand sides of lexical rules and application of lexical rules used in the derivation.

```
(2)
```

[[bowl],[Brow],[hamsymmlr,hamflathand,hamfingerbendmod,hamextfingeror,hampalmul,hamtouch, hamparbegin,hammoveur,hamsmallmod,hamarcd,hamreplace,hampalml,hamparend,R]] ---> PHON,SYN,SEM

We have adapted the left hand side (LHS) of ALE lexical items to be a list structure consisting of a sign gloss (an English description of the sign) and lists of HamNoSys symbols which may include variables for non-manual and manual sign features. Successful generation produces a list of signs (each of the latter containing a list of its phonology). Example (2) illustrates a typical lexical entry, here 'bowl'. However, the use of variables in the LHS prohibits use of ALE's lexical rules to characterise phonological relationships. Lexical rules are applied during lexicon compilation, deriving new lexical entries (e.g. for plurals) from existing ones (e.g. singular forms). During the lexical compilation process the LHS and RHS of lexical entries are treated as independent, hence our use of variables in the LHS of lexical entries results in the loss of the association with variables in the RHS for the plural form. However, via unification and using principles, it is possible to instantiate the phonetic structure (PHON) on the right hand side, and propagate this to the LHS using a plural principle. In example (2), the unistantiated non-manual (eye-)Brow movement that accompanies the manual features of the sign is determined by the mode of the sentence via the first non-chain rule, which associates this semantic input feature with the phonological (eye-)Brow movement. The variable R is used for distinguishing singular and plural forms for the sign (see below). This solution has the positive side effect, that a dynamic lexicon is created without increasing compilation time.

Currently, the implementation employs a relatively small lexicon for 50 signs (mainly from a kitchen domain). However, these entries contain a variety of challenging verbs, nouns, pronouns and modifiers which permit investigation of significant sign language grammatical constructs.

The current ALE implementation has 8 rules for BSL. These rules deal with sign order of (pre-/post-)modifiers (adjuncts) and (pre-/post-)complements. Rules of the standard HPSG model, which were designed mainly for English, have been modified to reflect the character of sign languages. BSL is a topic-comment language (mainly but not necessarily SOV), hence a SUBJ, COMPLEMENT distinction is less appropriate. In the SYN component of a lexical entry, PRECOMPS and POSTCOMPS features permit it to subcategorize for its own kinds of complements. From this follows the introduction of recursive precomp- and postcomp-rules which permit an arbitrary number of complements. To compensate for the lack of a Subject-Head rule or schema, a terminating rule - the Last-Complement rule - has been introduced. The last complement is therefore not necessarily the subject. The subject is just one of the complements,



Figure 2: Precomps and Postcomps rules

which can be identified by feature-sharing between the lists of complements and the SEM substructure (see example (1), where Indgreek1 = arg). One of the rules also changes the usual sign order if the English sentence subject is a pronoun so that the pronominal sign occupies sentence final position.

### 3.5 Principles

Currently there are 4 kinds of principles, which deal with sentence mode, pluralization of nouns and verbs, subject and object pronoun drop.

The mode principle inspects the MODE feature in the semantic component and returns a value for the facial expression which has to accompany the signing. In Example (1) the mode of the sentence is declarative (MODE:decl), therefore the feature of BROW is instantiated to a neutral expression, which is *non\_raised*. Brows have to be *furrowed* or *raised* in wh-questions and yes-no questions respectively. Other non-manual components, such as mouthing and body posture are yet to be encoded.

In BSL nominal plurals can be expressed in several different ways. Some plurals are formed by repeating the sign (usually three times) each repetition beginning at the location where the previous finished. Neither singular signs which involve internal repetition nor body anchored signs (ones which involve contact between the dominant hand and another part of the body) can be pluralized in this way. However, such signs can take a proform (a 'pronominal' handshape classifier) which can be repeated in a comparable fashion. Quantifiers, which occur before the noun in BSL, are also used for pluralization, but quantifiers can also be expressed as part of the internal construction of a sign. The distributive movement of the verb expresses that members of a group are involved individually in an action, but sweeping movement indicates the collective involvement of the whole group (distinguished in the SYN:HEAD:AGR:NUM:COLLORDIST feature). Repetition of some verbs can mean either that one individual repeats the action or that many individuals perform the same action (Sutton-Spence & Woll 1999). Of these possibilities for noun pluralization, nouns which can be repeated (by instantiation of the variable R in Example (2) to the HamNoSys symbol ham repeat continues everal) and non-repeatable ones with external quantifiers are handled. However, for pluralization of the remaining group of nouns the feature structure design contains relevant classifier information about the possible proforms (substitutors) for future development.

The plural\_principle for nouns takes the SEM:COUNT:NUMBER information from the SEM component (COUNT:NUMBER in Example (3)). The lexical item determines whether it allows plural repetition, if so, then the PHON:MAN:MOV:REPEAT feature (MOV = movement) is instantiated to the HamNoSys symbol expressing repetition in different locations (hamrepeatcontinueseveral) and COUNT from the semantic input is propagated as the current value to the noun's SYN:AGR:NUM feature while PHON:MAN:MOV:REPEAT is instantiated to 'no' to prevent further pluralisation. In all other cases PHON:MAN:MOV:REPEAT remains uninstantiated. This REPEAT feature in PHON is associated with the same variable as R from the lexical item LHS above.

#### (3)

Verb pluralization is handled in a similar way, however the repeated verb motion is only permitted if the index of the semantic role and the index of the appropriate complement are identical.

Sign languages typically contain verbal signs which allow pronoun drop (prodrop), where one or more of the subject, object or indirect object (or actor, theme, addressee respectively) are omitted and incorporated within the sign for the verb itself. The actor and addressee are included within the sign for the verb as starting and/or end position of the movement (so-called directional verbs). In the case of a direct object pronoun the handshape of the sign for the verb is inherited from the object/theme (so called classifier proforms). This phenomenon reflects a similar relation between rich agreement and non-overt expression of subject/object pronomina (Bos 1993), as in many languages, such as Italian and Hungarian. Indeed prodrop is found also in Chinese, which allows for 'topic-drop' without such rich morphology. Topic-drop is also possible in sign languages, but this is not currently addressed in our grammar.

The non-overt realization of the pronomina (prodrop) is catered for by an empty lexical entry whose LHS is instantiated to an empty list and has non-instantiated RHS feature structure values<sup>3</sup>. When the complement rules are processed, the prodrop principles check the semantic head for the values of subject and object prodrop features in all three persons. The possible values are *can*, *can't* and *must*. If it is *must*, an empty

 $<sup>^{3}</sup>$ ALE supports empty categories, however they could not been used to our purposes of prodrop. Empty categories are declared as lexical entries in a special format, therefore they suffer from a similar deficiency as lexical rules (see Section 3.4). Inheriting syntactic information from another lexical entry would not be possible in this way without duplicating lexical entries and therefore increasing the size of the static lexicon.

lexical item is chosen (of type *dropped* rather than type *word* to avoid ambiguity)<sup>4</sup>. The feature structure for such a lexical item is looked up in the lexicon using the RELN feature in the SEM component, (in fact to achieve this we drop out of ALE into in-line Prolog and use its ALE representation of lexical entries using the ALE operator (—>). In this way, the required SYN information of the empty string, which has to be unified with the complement information of the verb, is instantiated. This is an important step, as the verb may need the index of the noun for start and end position of the prodrop value is *can't*, generation proceeds normally, generating the daughter in the usual way for separate signs. If the value is *can*, both solutions are generated, however a preferred order is realized by arranging the order of prolog predicates accordingly.

# 4 Current Status and Conclusions

Currently the translation system of CMU linkages into the DRS-based intermediate semantic representation handles approximately the most common 50% of CMU link types. The following linguistic phenomena are included within this: transitive, intransitive verbs, temporal auxiliaries, passive, imperative, an unrestricted 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.

The HPSG based synthesis sub-component involves a small sign lexicon but with a sufficient variety of different kinds of signs to allow us to explore the use of constraint based unification for sign language generation. Specification of HamNoSys descriptions of sign requires attention to detail but resources exist for deriving this information. Formulation of rules of sign sequence formation, however, is more complex. The initial indications are that despite some technicalities which have had to be overcome in using ALE as an implementation platform, this is a fruitful approach for testing initial hypotheses of sign language grammar.

### References

Babelfish:2001, http://world.altavista.com

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

Brien, D. (Ed.): 1992, Dictionary of British Sign Language/English. London, Boston.

- Bos,H.: 1993, Agreement and Prodrop in Sign Language of the Netherlands. In: Drykoningen,F, Hengeveld,K (Ed.):Linguistics in the Netherlands. Amsterdam/Philadelphia.
- Bos, J., Mastenbroek, E., McGlashan, S., Millies, S., Pinkal, M.: 1994, A Compositional DRS-based Formalism for NLP Applications. Report 59. Universitate des Saarlandes.

Carpenter, B., Penn, G.: 1999, The Attribute Logic Engine. User's Guide. Version 3.2 Beta. Bell Labs.

Copestake, A., Flickinger, D., Malouf, R., Riehemann, S., Sag, I.: 1995, Translation Using Minimal Recursion Semanitics. In: Proceedings of the 16th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95).

 $<sup>^{4}</sup>$ Ale's empty category could not be used for this purpose for the same reasons as lexical rules discussed above.

- Copestake, A., Carrol, J., Malouf, R., Oepen, S., and others: 1999, *The (new) LKB System*. Version 5.2. Documentation.
- Cormier, K.A.: 1998, *Grammatical and Anaphoric Agreement in American Sign Language*. Graduate School of the University of Texas at Austin. Master Thesis.
- Cox,S.J., Lincoln,M., Nakisa,M.J., Coy,J.: 2001, The Development and Evaluation of a Speech to Sign Translation System to Assist Transactions. In: Submitted to International Journal of Human Computer Studies.
- Dorna, M., Emele, M.: 1996, *Semantic-based Transfer*. Report 143. In: Proceedings of the 16th International Conference on Computational Linguistics (COLING-96). Copenhagen.
- Dorna, M.: 2000, *Semantikbasierter Transfer*. (Doctoral Thesis). In: Arbeitspapiere des Instituts fuer maschinelle Sprachverarbeitung (AIMS). Vol.6, No.2.
- Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I.: 2000, *The Development of Language Processing Support* for the ViSiCAST Project. In: Assets 2000. 4th International ACM SIGCAPH Conference on Assistive Technologies. New York.
- Hennessy, J.L., Patterson, D.A.: 1996, Computer Architecture : A Quantitative Approach (2nd ed). Morgan Kaufman.
- Hutchins, W.J.: 2001, Machine translation and human translation: in competition or in complementation? In: International Journal of Translation, vol.13, no.1-2.
- Kamp,H., Reyle,U.: 1993, From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer Academic Publishers.
- Kennaway, J.R.: 2001, Synthetic Animation of Deaf Signing Gestures. In: The Fourth International Workshop on Gesture and Sign Language Interaction (GW2001). City University, London, UK.
- Klima E. and Bellugi U.: 1979, The signs of language. Harvard University Press.
- Kyle J.G., Woll B.: 1985, Sign Language The Study of Deaf People and their Language. CUP.
- Marshall, I., Safar, E.: 2001, Extraction of Semantic Representations from Syntactic CMU Link Grammar linkages. In: Recent Advances in Natural Language Processing (RANLP), G. Angelova et al (ed), Tzigov Chark Bulgaria.
- Neidle C, Kegl J, MacLaughlin D, Bahan B, Lee R.G.: 2000, *The Syntax of American Sign Language*. MIT Press.
- Pollard, C., Sag, I.A.: 1994, Head-Driven Phrase Structure Grammar. The University of Chicago Press, Chicago.
- Prillwitz,S., Leven,R., Zienert,H., Hanke,T., Henning,J., others: 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.
- Safar,E., Marshall,I.: 2001, Translation of English Text to a DRS-based Sign Language Oriented Semantic Representation. In: Conference sur le Traitment Automatique des Language Naturelles (TALN) vol 2, pp297-306.
- Shieber, M., van Noord, G., Moore, C., Pereira, F.C.N.: 1989, A Semantic-head-driven Generation Algorithm for Unification-based Formalisms. In: 27th Annual Meeting of the Association for Computational Linguistics. Vancouver. pp 7-17.
- Sleator, D., Temperley, D.: 1991, Parsing English with a Link Grammar. Carnegie Mellon University Computer Science technical report CMU-CS-91-196.
- Sutton-Spence, R., Woll, B.: 1999, The Linguistics of British Sign Language. An Introduction. University Press, Cambridge.
- Wideman, C.J.: 2000, Internet-enabled, 3D, SigningAvatar Software Offers Accessility for the Deaf. http://www.signingavatar.com/nc/presentations/april\_fose.html.
- Woll, B.: Language, culture and identity: insights from sign language and the deaf community. In: The Linguist, August/September 2001.