ViSiCAST Deliverable D5-3: Proto-text-to-sign notation
Peer Review

Project Number:	IST-1999-10500
Project Title:	ViSiCAST Virtual Signing: Capture, Animation, Storage and Transmission
Document Type:	Peer Review

Deliverable Number:	**D5-3**
Contractual Date of Delivery:	June 2001
Actual Date of Delivery:	August 2001
Title of Deliverable:	Proto-text-to-sign notation
Work-Package contributing to the Deliverable:	Workpackage 5 (Language and Notation)
Nature of the Deliverable:	PR (Prototype)
Author(s):	Ralph Elliott, Thomas Hanke, Ian Marshall, Constance Schmalling, Eva Safar, Margriet Verlinden

Peer Review Author(s):	Kurt Eberle
Date of Review:	November 2001

Qualifications and Expertise as Peer Reviewer:

I studied romance languages and mathematics at the universities of Tübingen, Freiburg and Paris Sorbonne and computational linguistics at the university of Heidelberg. I completed my studies with 1st and 2nd Staatsexamen romance languages (subject french) and mathematics (1983/1985), with the diploma in mathematics (1987) and a doctor's diploma in philosophy tutored by Hans Kamp (1991).

From 1987 until 1999 I held research positions at the Institut für Maschinelle Sprachverarbeitung at the University of Stuttgart and at the Scientific Center of IBM in Heidelberg, where I did research and teaching mainly about semantic subjects (DRT), logics and machine translation.
At IMS and IBM I was responsible for several scientific projects including transfer of VERBMOBIL.
Since 1999 I am shareholder and head of 'development french' and 'development transfer technologies' at linguatec Entwicklung & Services.

x	**I have read the report of the Deliverable.**
x	**I have read the Workpackage Description relating to the Deliverable.**

Rating of the report as a whole:

1	2	3	4	5	Poor = 1, Excellent = 5
			x		**Overall quality**
			x		**Reflecting the state-of-the-art**
				x	**Meeting the objectives of the Workpackage**
			x		**Meeting a real need**
			x		**Contributing to this field**
			x		**Style and clarity of the report**

Executive Summary:

The deliverable describes the architecture of a system which translates English text into terms of a sign language based upon the HamNoSys format.

With the deliverable comes a demo system which corresponds to this architecture. It produces correct syntactic structures for the sample sentences (CMU linking structures), from which lambda-DRSs are abstracted which are incorporated into a text representation in the sense that their discourse referents are related to each other by anaphora resolution and that the new sentence is evaluated w.r.t. pragmatic structure und function. These augmented sentence representations are translated into terms of the HamNoSys sign language. Unfortunately, there is not yet a visualization component which could help to understand the expressions and actions of this sign language.

The paper describes the architecture of the system and the different components sufficiently in order to be able to understand the syntactic and semantic output structures of the demo. It gives a good understanding about the scenario and the attainable coverage with respect to vocabulary and syntactic and semantic structures, i.e. to expressivity.

It has been good strategy, so it seems, to make use of existing components and algorithms respectively, which are: the CMU parser, lambda-DRS based semantic construction, MRS/HPSG based generation (Lingo) and HPSG SEM based generation respectively (ALE). The choice of the different components is motivated in the paper, though one would like to find a more detailed discussion about the reasons for preferring HPSG to LFG (in order to take up a common debate) and lambda-DRT to other types of DRT and so forth.

The system is hybrid in the sense that for analysis the CMU-linkage grammar is used whereas for generation (of signs) an HPSG approach is pursued. (This is true for both system variants: the Ale- and the Lingo-type). Also, at the level of semantic representation, the analysis outputs DRSs which must be mapped onto MRS structures (or onto the semantic structures of ALE respectively) which are the (semantic) input of generation.

In the presence of existing systems which are more homomorphic in this respect (think of initiatives of Xerox and others for parallel LFG grammar development or Verbmobil's HPSG approach and others), it would have been fair to say more about the motivation of this. In any case it should have been said something about the problems which certainly arise when mapping the different formats onto each other. Also, with respect to the foundations of the approach as such, more comments about anaphora resolution and about the strategies used there, in particular with respect to accessibility in underspecified structures (unresolved lambda-DRSs) would have been very welcome.

Another question arises with regard to extendibility (not practical extendibility in the first place, but theoretical extendibility). This means to reflect upon the question whether the theoretical basis is chosen such that, potentially, it can provide means for representing and processing richer communication, which uses more complex language and dialogue features. We emphasize that partly this is done, however statements about phenomena like modal operators and the like w.r.t. vocabulary and about structural ambiguity would have been welcome also.

The paper certainly presents sound work. It is intelligibly written and mostly gives the right examples and comments such that the reader can get a feeling for the scenario and the attainable coverage of the project.

Comments on Contribution to objectives of Workpackage and ViSiCAST project:

1) The task
The claim of the paper is to demonstrate the/a provisional route from English text to sign language notation using the sign coding suggestions HamNoSys and SiGML. This conforms to the objectives of workpackage 5 as described in the VisiCAST proposal, annex 1-description of work and, actually, is carried out. The visualization of the translation, this means working out the animation component which translates the sign expressions into gestures and movements of an avatar is part of workpackage 4.

2) The motivation
The VisiCAST proposal aims at translating speech and text respectively into avatar gestures in order to provide support for deaf people with respect to the considered fields of application, which are (simultaneous ?) translation of TV information into signing, corresponding web page translation and the real world scenario of a post office. The relevance of translating speech is immediate. The relevance of translating text is motivated by the fact that for most deaf people, as it seems, all types of verbal skills are difficult to acquire, including reading. It is obvious that the opposite direction is of equal relevance (because enabling understanding the signing of deaf people), though presumably being much more complex. Nevertheless the investigation, the architectural design and the implementation should take into account this possible future objective. This means, one should attach importance on discussing things from the prespective of reversibility. This is done with respect to motivating the choice of HPSG for generation, but this is missing with respect to considerations about the overall architecture and the remaining modules. Also one should be more conscious of the fact that the scenarios of translating speech and translating text respectively are rather different with respect to expectations about run time perfomance (whereas with speech applications, normally, real time performance is expected, with regard to text it is not). Also they are different with respect to phenomena of textual coherence and especially with respect to types of referring. The latter at least has an immediate effect on the work of workpackage 5 and some comments about this in the deliverable would be appreciated.

3) The domain and its representation
The paper reports that the considered scenario provides less than 500 words, that, however, some representatives of the relevant class of the so called 'classifier verbs' are included and that different types of sentence mood, plural and locative phrases are dealt with, such that sufficiently complex requests and communications about the objects of the chosen 'multi agent kitchen scenario' can be represented (which appears to be a modern variant of the well known 'blocks world' with a signing self and a co-agent and which, of course, fits with the post office scenario of the ViSiCAST proposal). All this demonstrates sound judgement about what is feasible on the basis of the available state of the art and the given constraints about time and man power.

Feature structure definitions are presented for the relevant sign types such that the mapping between SEM- or MRS-information into (gesture-)PHON-information via HPSG-signs and the corresponding principles and rules can be illustrated sufficiently. Nevertheless, additional information about what signer's equivalents are of logically complex verbal elements like modal operators, like focus adverbs, like discourse particles and other material of this type, about how these means could be represented in the formal sign languages (HamNoSys /SiGML) and how the mapping could look like, would have been very welcome, in order to give a feeling whether a system with broader coverage could be built from the suggested prototype.

4) The architecture
It is praiseworthy that the Visicast project tries to build upon existing components in order to concentrate on the relevant contributions which one expects from this type of project -- which must be to obtain more precise formal grammars of sign languages with broader coverage and to spell out the transfer components which use them for generation (and analysis). The pre-defined components used in D5-3 are the CMU linkage parser and the HPSG-generation systems Lingo, for the generation of HamNoSys-expressions of DGS, and Ale, for the generation of HamNoSys-expressions of BSL and NGT. In addition, one relies on a variant of underspecified DRT, which grosso modo is the DRT version underlying the Verbmobil system and which provides the interface formalism between the CMU-parser output and the semantic representations which are the input of the Lingo and the Ale system respectively. Nevertheless, comments which motivate the choice of exactly these components would have been welcome. Between in- and output of these components

and the visualization of workpackage 4 respectively the following interface modules and formalisms are defined (and demonstrated by the demo system):

a) The mapping of the CMU-linkage analyses of english sentences to the expressions of the used underspecified DRT, which are the so-called lambda-DRSs;

b) the translation of lambda-DRSs into the domain of the SEM-expressions of the HPSG-signs used in Lingo and Ale respectively and

c) the definition of the basic features of a sign grammar which define the shape that the SYN-values of the Lingo and Ale HPSG-signs can take, together with the corresponding principles and compositionality. This definition is accompanied by projections of the basic material to phonological information, which, here, is expressed in terms of HamNoSys descriptions.

This carries out the D5-3-relevant objectives of the ViSiCAST proposal to a satisfactory extent: As it seems, the fragment planned for this stage is not completely achieved but to a sufficiently high degree, comprising the most relevant structures.

5) General Remarks
Some general remarks seem to be appropriate w.r.t. the overall setting.

a) Parallel development of modules/formalisms
Apparently, the sign grammar is developed simultaneously with the other modules, in particular with the DRS-analysis of text and speech and with the different transfer modules. Since developping formal grammars for sign languages is relatively new ground, translatability betwen semantic representations of text and speech and representations of sign language is an interesting question: There might be verbal statements which cannot be (directly) translated into sign language and vice versa. Thus, it would be good strategy to assume some relatively broad fragment of speech and text on the one hand and of sign gestures on the other in order to carefully define the maximal logical expressivity the translation scenario may take (source, target and
transfer relation), such that, later, when working out step by step the different phenomena it cannot happen that one of the grammar presupposes some additional formal means which (modulo translation) exceed the expressivity of the formalisms with which it communicates.
In short, specification of the (envisaged) coverage of the different component formalisms should preceed the working out and implementiation of the components.

b) Context, underspecification (lambda-DRSs) and anaphora resolution
The architecture is completely sequential. This means that subsequent processing steps cannot provide feed-back for earlier steps. It is reasonable -- for this reason and for reasons of robustness also -- as is planned, to provide possibilities for human interaction (in order to apply corrections and to choose between competing solutions). This means that the context can be adjusted manually, if needed.
According to the paper, at present, the generation (does it mean the processing?) is restricted to single sentences. If this is true, contextual knowledge is restricted to sentential information. It is known that with respect to resolving pronouns and descriptions in verbal utterances, it makes a great difference whether a sentence can be assumed to be the first (or only) of a discourse or whether it completes preceeding information. (For instance, cataphora are more easily acceptable in single sentence contexts than in others etc.). Thus, extending the system in this respect means providing a more complex computation of reference with revised preference weighting. Also, it is known that speech and text behave differently with respect to referring (different use of indexicals and the like). Therefore, in order to correctly deal with an incoming analysis, it has to be taken into account whether it stems from speech recognitation or from a text. Using lambda-DRT for representing source information allows for underspecified representations, at least potentially (though, as it seems, at present, the lambda-terms of the sentence are resolved into one specified representation). If, in order to avoid overdeterminism with respect to the task of translating, one sticks to the underspecified format, one has to be aware of the fact that resolving pronouns on this basis is a relatively new research subject, such that there aren't enough reliable rules of resolution yet. Also the expressivity of lambda-DRT and MRS is not the same (if underspecification really is exploited).

c) Components
Besides the asymmetry of the semantic source and target representation and syntactic parsing and generation, from the perspective of architecture, the question also arises why the generation grammar computes HamNoSys expressions and not directly SiGML expressions, which as it seems, are the input of the animation tools of workpackage 4.

Summarizing, in my opinion the objectives of the proposal with respect to D5-3 could be achieved satisfactorily. Comments would have been welcome however, which could provide some clarification about the questions raised.

Signed:	**Date: November 2nd 2001**