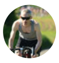
Linguistic modelling and language-processing technologies for Avatar-based sign language presentation

Article *in* Universal Access in the Information Society · February 2008

DOI: 10.1007/s10209-007-0102-z · Source: DBLP

John Glauert
University of East Anglia



Richard Kennaway
John Innes Centre

Univ Access Inf Soc (2008) 6:375–391
DOI 10.1007/s10209-007-0102-z

LONG PAPER

Linguistic modelling and language-processing technologies for Avatar-based sign language presentation

R. Elliott · J. R. W. Glauert · J. R. Kennaway · I. Marshall · E. Safar

Published online: 26 October 2007

Abstract Sign languages are the native languages for many pre-lingually deaf people and must be treated as genuine natural languages worthy of academic study in their own right. For such pre-lingually deaf, whose familiarity with their local spoken language is that of a second language learner, written text is much less useful than is commonly thought. This paper presents research into sign language generation from English text at the University of East Anglia that has involved sign language grammar development to support synthesis and visual realisation of sign language by a virtual human avatar. One strand of research in the ViSiCAST and eSIGN projects has concentrated on the generation in real time of sign language performance by a virtual human (avatar) given a phonetic-level description of the required sign sequence. A second strand has explored generation of such a phonetic description from English text. The utility of the conducted research is illustrated in the context of sign language synthesis by a preliminary consideration of plurality and placement within a grammar for British Sign Language (BSL). Finally, ways in which the animation generation subsystem has been used to develop signed content on public sector Web sites are also illustrated.

1 Introduction and background

1.1 Overview

This paper presents an account of work on the development of techniques and software tools to support the generation of natural sign language performances by a computer-generated virtual human, or *avatar*. Much of this work has been undertaken as part of the ViSiCAST project, over the period 2000–2002, and more recently in the eSIGN project, 2002–2004. Both these projects, which received substantial funding from the European Union under the 5th Framework Programme, have been concerned with the development of technology and standards to support communication for and with deaf people in sign language using avatars.

The application areas considered for this technology were broadcasting, interactive Internet information, and support for communication between counter clerks and deaf clients for retail and public information provision [10]. Examples of these applications will be provided later on in this paper.

As explained in more detail below, a cardinal precept in these projects has been that the technology they develop should respect as fully as possible the distinctive linguistic characteristics of sign language.

A major strand of activity in the ViSiCAST project was the development of a prototype system to demonstrate the feasibility in principle of starting with a text, representing spoken language material, and generating from this an equivalent sign language performance by an avatar. It is this text-to-sign-language-animation system, and various further developments and applications of it, that are reported here. The basic approach taken in the development of this system was outlined, in advance of its actual

R. Elliott · J. R. W. Glauert (✉) · J. R. Kennaway ·
I. Marshall · E. Safar
School of Computing Sciences,
University of East Anglia,
Norwich NR4 7TJ, UK
e-mail: jrwg@cmp.uea.ac.uk

development, in [7]. A crucial feature of this approach is the decomposition of the system into a sequence of processing stages, each applying a transformation at a particular level of linguistic analysis to the input it obtains from the previous stage, and passing the result of this transformation to the following stage; broadly speaking, earlier stages correspond to higher levels of linguistic analysis, whereas later stages correspond to lower levels.

The most prominent of the interfaces between pairs of adjacent stages in this processing pipeline is the phonetic-level interface, which comes more or less in the middle of the pipeline. The decision was taken at the outset to base this interface on the Hamburg Notation System, generally known as *HamNoSys* [13, 25], a well-established notation originally developed by the University of Hamburg for the transcription of sign languages at the phonetic level. The term "phonetic" here referring to the appropriate features of manual and non-manual sign language articulation, as opposed to the vocal articulation used for a spoken language. This phonetic-level interface based on HamNoSys has given a stable reference point for much of the work described here. One part of this work has consisted simply in the definition and refinement of the interface itself, and in particular in the definition of the SiGML (Signing Gesture Markup Language) notation, an XML application based on the HamNoSys phonetic model. The remaining work on the ViSiCAST text-to-signed-animation system then divides naturally into the development of two complementary subsystems:

- A *front-end*, dealing largely with issues at a higher level of linguistic analysis, i.e., with syntactic, semantic, and morphological issues, with reference both to the given spoken language and the target sign language; the output of this subsystem being a description of a sign language utterance at the phonetic level (expressed either in HamNoSys or SiGML).
- A *back-end*, which in linguistic terms operates from the phonetic level down, whose input is a HamNoSys/SiGML description of a sign language utterance, and whose function effectively is to treat this description as a "script" from which the corresponding animation is generated by an avatar on the screen of a computer or a mobile device.

The structure of the paper, which is based on the architecture of the software system as just described, is as follows. The remainder of this introductory section provides a brief review of the authors' earlier experience using motion-captured data to drive a signing avatar, an outline of the motivation for the development of the ViSiCAST text-to-sign-language-animation system, and a brief outline of the HamNoSys notation which, as previously explained, is an important reference point for a more detailed understanding of the system. Section 2 discusses the front-end subsystem in some detail, describing the translation system architecture and the kinds of linguistic issues that arise in this context, focusing particularly on the mismatch between plurality in English and British Sign Language (BSL), and on the interaction with placement in sign languages. Section 3 gives further details on the organisation of the back-end subsystem which generates the appropriate virtual human animation from a given HamNoSys or SiGML description of a sign language utterance. Section 4 describes alternative techniques for dynamic generation of signed content via HamNoSys/SiGML descriptions, that is, techniques not based on translation of English text as described in Sect. 2; Sect. 4 also describes some of the web-sites and other applications developed using these techniques by other members of the eSIGN project. Finally, Sect. 5 provides conclusions.

1.2 Natural sign languages and animation based on motion capture

During the past half-century, sign languages have been recognised as genuine minority languages which coexist with majority languages [23] and which are the native languages for many deaf people. There is some ambiguity in the term *sign language*, which is sometimes taken to cover a spectrum which includes (abbreviated) transliterated forms of spoken languages (Sign Supported variants), as well as genuine sign languages with their own distinctive linguistic characteristics. However, the former are really artificial forms of sign language, often used to educate and inform deaf people via the geographically local spoken language rather than a preferred form of natural communication. In this paper, the term *sign language* is used to denote natural sign languages.

Provision of information access and services in signed languages is as important as for other minority languages. Such provision, however, introduces theoretical and technical challenges. Initially, the conducted research was motivated by an exploration of the utility of motion-captured data to present Sign Supported sign language variants, while investigating ways in which avatars might be used to increase the volume of signed accompaniment to broadcasting [24]. This approach involved the use of motion-capture techniques to record the performance of individual signs by a (real) human signer, allowing a lexicon of signs to be created. Sign sequences were then dynamically constructed from this pre-recorded lexicon based upon the transliteration of an English text. The main advantage of this approach over pre-recorded video clips of human signing is that it supports the generation of smooth signing sequences by means of digitally generated blending from the end of one sign to the start of the next. Nevertheless, it was clear from this work that the approach based on motion capture suffers from a severe limitation, in that it does not permit decomposition of a sign into a number of parallel activities which are inherent in genuine signing. At the coarsest level, virtual human signing generated through motion capture couples together manual gestures and facial expressions in a way that makes it difficult to re-use either component in a new context. At a finer-grained level, it is clear that some manual signs are constructed from separate phonetic components which are brought together in parallel to realise the complete sign. Thus, this early experimentation determined the need to address genuine sign languages as the target form of expression, and exposed the limitations of motion-captured data in furthering that end.

1.3 Role of sign language in the ViSiCAST project

It is often assumed that the increased use in modern life of computer-based technology effectively dispels the barriers to communication experienced by deaf people. Computer-based communication, so the reasoning goes, is predominantly visual and must therefore suit deaf people, whose visual capabilities are usually no different from those of hearing people. This argument fails to take account of the vital role of language in human communication. For many deaf people, notably pre-lingually deaf people, their first language is a sign language, rather than any spoken language. Written text, however, always represents some spoken language, and is thus “foreign” to many deaf people.

In fact, there are several layers to this foreignness. Firstly, it is important to appreciate that, although a spoken language typically has a written form with which all educated users become progressively more familiar from a relatively early stage in their childhoods, a sign language by contrast does not have an associated writing system at all. Hence, for deaf people the very medium of text is not as tightly bound to language use as is typical for hearing people. Moreover, when a deaf sign language user is confronted with the textual form of a spoken language, the language does not represent that deaf person’s native language, but the fact that the language itself uses an entirely different mode of articulation means that it exhibits linguistic characteristics that differ significantly from those of any sign language.

A cardinal guiding principle in ViSiCAST, motivated in part by previous experience with the motion-capture based system described above in Sect. 1.2, was that sign language is the natural and hence preferred means of communication for many deaf people, especially pre-lingually deaf people; and that in consequence, the technology developed by the project should support authentic sign language communication. In other words, the avowed intention was to develop technology which respects the character of sign languages as languages in their own right, recognising that they have their own distinctive characteristics at all levels of linguistic analysis. This principle has had a particularly strong influence on work on the ViSiCAST text-to-sign-language-animation system.

1.4 HamNoSys and SiGML

As explained in Sect. 1.1, the interface between the two parts of the ViSiCAST text-to-animation system is a phonetic-level description of the sign sequences to be performed by the avatar, using the Hamburg Notation System (HamNoSys) and the Signing Gesture Markup Language (SiGML).

It is a common misapprehension that sign languages are restricted in communicating information through the form and motion of the hands. However, sign languages involve the simultaneous use of manual and non-manual components in signed communication. Non-manual features consist of the posture of the upper torso, the orientation of the head, and facial expression. Manual features have often been decomposed as hand-shape, hand orientation, hand position and motion [3, 32, 34]. HamNoSys [13, 25] is the notation system developed by the University of Hamburg as a vehicle for the transcription of human sign language performance. The notation uses its own character repertoire, containing over 200 iconic symbols, for which a computer font is available. It effectively embodies a model of sign language phonetics that is sufficiently general to cover all sign languages. HamNoSys deals mainly, but not exclusively, with the manual aspects of signing, that is with the shape, position, orientation, and movement of the signer’s hands. To support the work of the ViSiCAST project a new version of the notation, HamNoSys 4, was developed: this introduced a few new features into the manual part of the notation, and considerably enhanced the part dealing with linguistically significant movements of the signer’s face, including eye-gaze, eye-brows, nose, cheeks, and mouth, as well as movements of the signer’s torso, shoulders, neck, and head.

The SiGML notation [5] is an XML application [2]. SiGML supports the description of sign language sequences in a manner suitable for performance by a signing avatar. Over time, SiGML has come to include several related but distinct forms of low-level sign language description, including an XML-compatible transliteration of HamNoSys itself. The central one of these forms of description is that established early in the ViSiCAST project as an interface notation for the text-to-sign-language-animation system, now sometimes referred specifically as *Gestural SiGML*. Gestural SiGML is based on HamNoSys 4 in the sense that its model of sign language phonetics is essentially that of HamNoSys, albeit with one or two generalisations, and it is compatible with HamNoSys in that any valid HamNoSys sequence has a fairly direct translation into Gestural SiGML.

To give the reader some feeling of the way the notation works, Fig. 1 shows a HamNoSys transcription for the DGS (German Sign Language) sign "HAUS" (house), together with a diagram illustrating how the signer's hands trace an outline first of the roof and then of the walls of a house specified by this transcription. In more detail, this sequence of HamNoSys symbols is interpreted as follows. The initial pair of dots specifies that the sign performance is symmetrical: everything specified subsequently for the signer's right (i.e., dominant) hand is to be mirrored by the left (i.e., non-dominant) hand. The next three symbols specify the initial hand posture: a flat hand with fingers pointing outwards, and the palm facing down and to the left. The following symbol specifies that the two hands are initially in contact. The remaining four symbols specify a sequence of three motions: firstly a movement downwards and to the right, secondly a change of orientation so that the palm faces to the left, and finally a movement downwards. This example illustrates several features common to almost all sign descriptions: the description of an initial configuration of the signer's hand, or hands, covering their shape, orientation, and location, followed by the description of one or more movements of the hands, where the term "movement" covers not only change of location, but also changes to one or more aspects of the initial posture (in this case, a change of hand orientation). The Gestural SiGML representation of this sign is shown later, in Sect. 4.

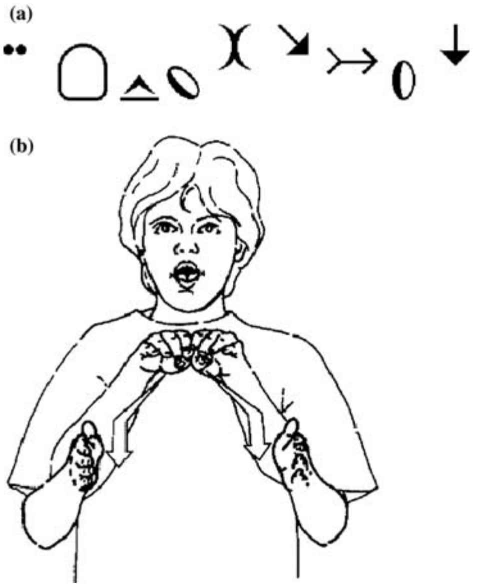


Fig. 1 HamNoSys notation, with illustration, for the DGS sign "HAUS"

2 Translation system architecture

2.1 Overview

The front-end of the ViSiCAST text-to-signed-animation system performs translation from spoken language to sign language. The system was designed to take English text as input, from which corresponding phonetic-level output can be produced for three distinct national sign languages, German Sign Language (DGS), Dutch Sign Language (NGT), and BSL, with one project partner in each country responsible for the generation of the corresponding sign language output. To meet these requirements, an innovative system was designed and implemented by researchers at UEA, together with co-workers at Viataal (formerly Instituut voor Doven) in the Netherlands, and in the IDGS (Institut für Deutsche Gebärdensprache) at the University of Hamburg. This system is described, with particular reference to the generation of BSL output, in [21, 22, 27–30].

The overall architecture of the translation system is shown in Fig. 2. Previous work has demonstrated the appropriateness of this approach for examples using singular nouns and verbs. A collection of automatic transformation components is potentially augmented by user interaction. English text is input to the CMU parser [31], which, assisted by lexical information from the WordNet database, outputs a set of links—a linkage—for a sentence. The CMU parser is robust and covers a high proportion of English linguistic phenomena. The parser often produces a number of linkages for one sentence. Currently, the user selects the correct linkage by direct intervention. The transformation from the appropriate output linkage to its Discourse Representation Structure (DRS) is performed using λ-DRS definitions associated with link types which are composed using λ-calculus β-reduction and DRS merging [1, 20].

The DRS representation [14] is used as a machine translation semantic transfer notation with a collection of transformations that convert the valencies of English-oriented predicates to more appropriate valencies for their BSL counterpart. The transformed DRS is converted into an Attribute Logic Engine (ALE) [4] Head-driven Phrase Structure Grammar (HPSG) [26] semantic feature structure, from which the HamNoSys phonetic form is generated using sign language HPSG-grammar rules.

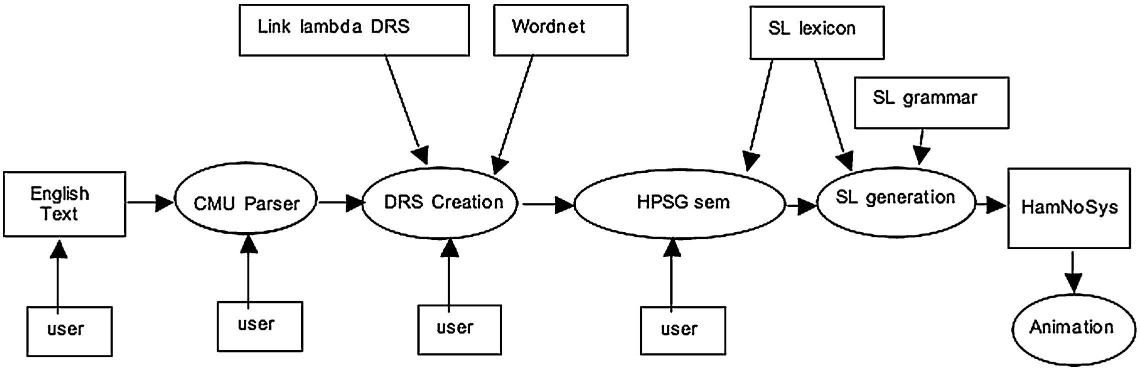


Fig. 2 The Translation System Architecture

Finally, the generated HamNoSys is given an executable interpretation by the avatar signing technology.

2.2 Placement in sign languages

The three-dimensional space in front of a signer extending from above head height to approximately the waist constitutes the *signing space* in which signs are articulated (see Fig. 3). A number of signs have a fixed form signed at a particular location, which is especially the case for body-anchored signs where the location at which the sign is articulated is part of its meaning, for example WOMAN, MUG and BUY in BSL. However, a number of nominal signs can be modulated (parameterised) by a specific location, for example, for BOWL and PLATE in BSL, the location at which these are signed can have significance. Furthermore, pointing (indexing) can be used to locate nouns at particular locations, and indeed nominals are often associated with a simpler handshape (proform) which can be used anaphorically and can be located at a particular location. Verbal signs can similarly have fixed (e.g., EAT in BSL) or parameterisable forms (e.g., TAKE in BSL). For such verbs, either the start or the end location or both may have syntactic significance. In summary [3, 34]:

- some nominals can be signed at specific positions in signing space, and these locations then have syntactic significance,
- nominals which cannot be located in this way can be positioned in signing space by indexing a particular location after the sign or by using the sign's classifier handshape (proform) and placing that at the location,
- nominals can be anaphorically referred by inclusion of classifier handshapes within manipulator verbs,
- directional verbs must be syntactically consistent with the locations of their subject and object,
- verbs exhibit syntactic agreement for number with their arguments.

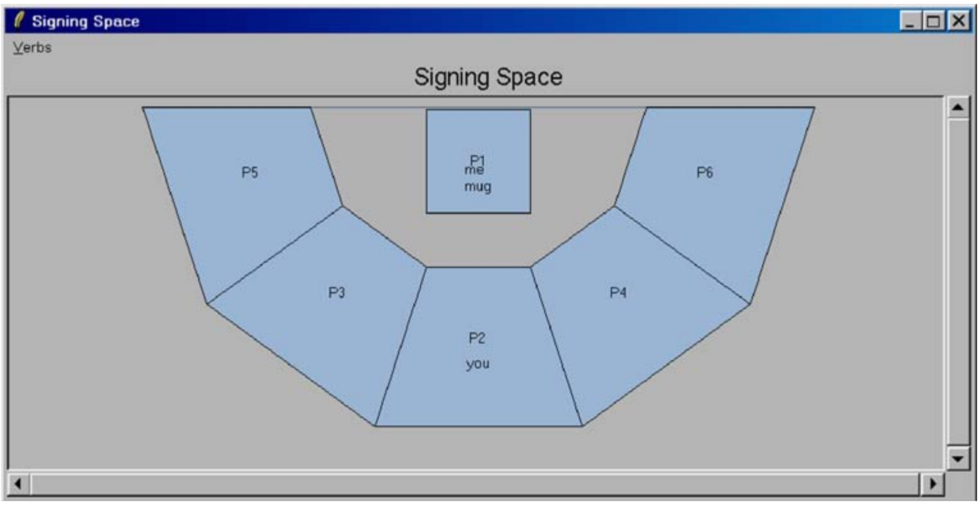


Fig. 3 Sign space allocation map (after communicating *I take the mug.*)

The level of granularity of signing space differs with the communicative context [19]. Some communications in sign are anchored to the physical environment in which they take place (so-called topographical uses of signing space) and location of signs and indexing is expected to be consistent with that environment. In such cases, signing space is considered to be sufficiently fine grained to be consistent with the environment. In cases where the objects and individuals are not physically present, the use of position within signing space is syntactically motivated and is considered to be more coarsely grained. In the adopted model, signing space is separated into five discrete "areas" at a specific height and focused around the position of the signer (see Fig. 3). Each of these positions can be specified in HamNoSys. The HPSG grammar generates sign sequences such that nominals are located in signing space by reference to an internal model of that space (the sign space allocation map), and directional verbs are parameterised by appropriate positions derived from it [22].

A consequence of this use of signing space is that particular positions in it can be populated by more than one object or person (e.g., having taken a mug both it and 'I' are situated at the same sign space location), although typically these can be distinguished by different classifier handshapes.

The sentence "*I take the mug.*" is glossed as

$$\mathrm{MUG}(px) - -\mathrm{TAKE}(px, p1, \mathrm{manip\_handshape}(\mathrm{MUG}) - -\mathrm{me}(p1)$$

where the original position of the 'mug' (px) and the position of 'I/me' (p1) must agree with the start and end positions of the sign for 'TAKE'. The fully instantiated HamNoSys phonetic form for the sign sequence generated from this sentence is:

```
[
[ [ mug ],      [ 'mVg', non_raised ],     [ ] ],
[ [ take ],     [ 'teIk', non_raised ],    [ ] ],
[ [ me ],       [ 'mi:', non_raised ],     [ ] ]
[ [ punct ],    [   ],                     [  ] ]
]
```

where each sign is a triple containing its English gloss, its non-manual components (a Sampa phonetic characterisation of mouthing and eye-brow position), and the HamNoSys manual description. Figure 4 shows the handshape and hand location (with arrows indicating the motion).

The synthesis of TAKE in this sign sequence is achieved, however, by a parameterised lexical entry for TAKE whose head is

```
[[take],['teIk', Brow],[ClassifierShape(orientedToSourcePosition),
                        SourcePosition,
                        R1,
                        h  ClassifierShape(orientedToDestPosition),
                        DestPosition,
                        R2]                        ] ---> RHS
```

During sign language generation, the HPSG grammar constrains the instantiated form to be consistent with the allocation map positions for MUG (SourcePosition and orientation) and 'I/me' (DestPosition and orientation) and the classifier handshape for MUG (ClassifierShape), thus instantiating these parameters to the specified HamNoSys subsequences for these component parts.

The overall organisation and management of signing space over time throughout a sign commentary (text) is a complex and under-researched topic. The remaining paper focuses on the possibilities for representing plural phenomena in sign language under both distributive and

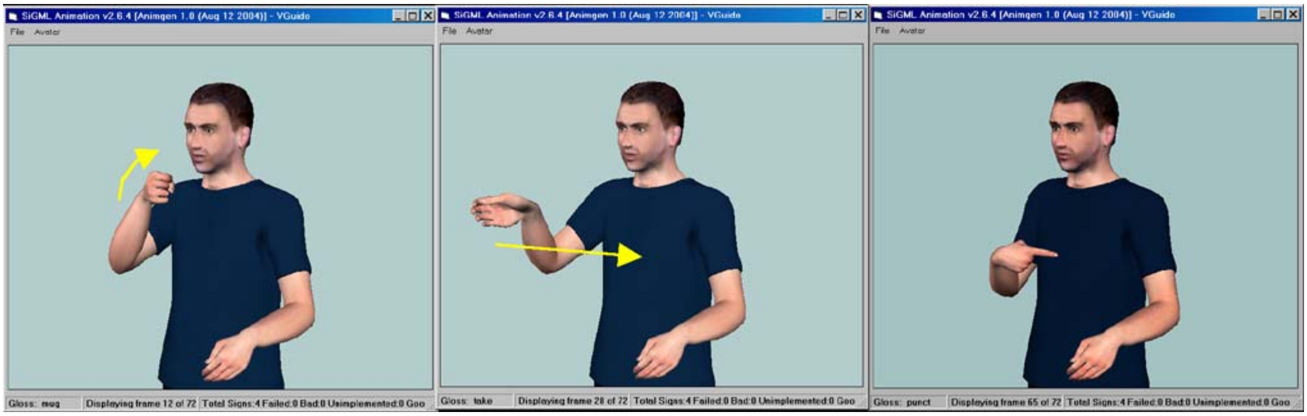
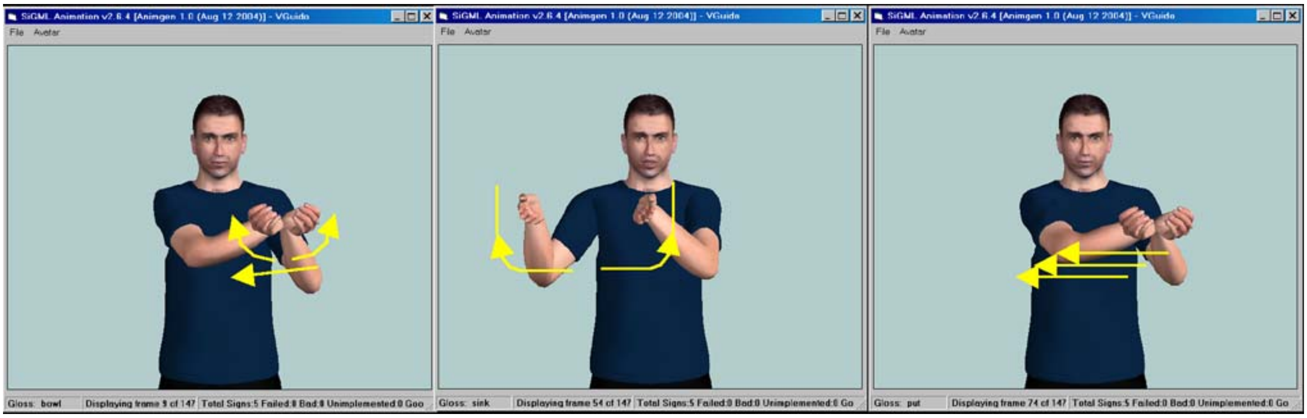

SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar
Gloss: mug
Displaying frame 12 of 72
Total Signs:4 Failed:0 Bad:0 Unimplemented:0 Goo
SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar
Gloss: take
Displaying frame 28 of 72
Total Signs:4 Failed:0 Bad:0 Unimplemented:0 Goo
SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar
Gloss: punct
Displaying frame 65 of 72
Total Signs:4 Failed:0 Bad:0 Unimplemented:0 Goo

Fig. 4 MUG TAKE I


SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar
SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar
SiGML Animation v2.6.4 [Animgen 1.0 (Aug 12 2004)] - VGuido
File Avatar

Gloss: bowl | Displaying frame 9 of 147 | Total Signs:5 Failed:0 Bad:0 Unimplemented:0 Goo

Gloss: sink | Displaying frame 54 of 147 | Total Signs:5 Failed:0 Bad:0 Unimplemented:0 Go

Gloss: put | Displaying frame 74 of 147 | Total Signs:5 Failed:0 Bad:0 Unimplemented:0 Go

Fig. 5 BOWLS SINK PUT (I)

collective interpretations, and discusses an approach to addressing the issues raised in an English-text-to-sign-language translation system.

2.3 Plurals in sign languages and HamNoSys

Sign languages provide a number of forms which can be used to characterise plurals. Signs for nominals can be numerically quantified (THREE MUGS) or can be modified by quantifying signs such as SOME, MANY, ALL. Many nominal signs can be pluralised by repeating the sign (or an abbreviated form of it), and such signs can often be located in signing space like their singular form. The way such plurals interact with the use of signing space for placement gives rise to interesting linguistic possibilities.

Sign languages, like spoken languages, make a semantic distinction between collective (the entire set) and distributive (each individual within the set) uses of plural nouns and verbs. In English the range of collective to distributive interpretations are realised in constructs which are more or less syntactically explicit in their preferred interpretation [14]. Specifically, definite and indefinite plurals are treated as having a preferred collective interpretation, whereas universally quantified and numerically quantified nominals are treated as having a distributive interpretation.

The girl lifted the books.	(collective)
The girl bought the teachers a present.	(collective)
I put the cups in the sink.	(collective)
The girl bought most teachers a present.	(distributive)
The girl bought three teachers a present.	(distributive)
I put three cups in the sink.	(distributive)

Nonetheless, for most collective interpretations a secondary interpretation exists and in context can be brought to the fore. Thus, definite and indefinite plurals are ambiguous out of context. Furthermore, some verbs appear to demand one reading rather than another, for example

I juggled three balls.	(collective)
The girls ate their dinner.	(distributive)

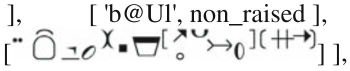
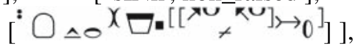
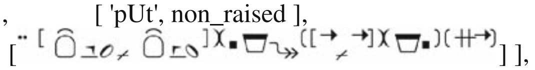
'Juggle' appears to require a collective interpretation ('juggling' a single ball on three different occasions hardly constitutes juggling). Conversely, 'ate' pragmatically implies a distributive interpretation even when the preference indicated syntactically is collective.

Irrespective of whether this analysis of which interpretation is triggered by which syntactic construct is correct or not, the following are noted. In English, some constructions are under-specified in the precise nature of the relationship involved, and still native speakers appear to cope with that. Furthermore, Discourse Representation Structures (DRSs) as a semantic formalism are able to represent the distinction between the interpretations.

In sign languages, however, for a number of these cases it is impossible to remain neutral with regard to whether a collective or distributive interpretation is intended. Thus, translation between English and BSL must address comparable issues to translations between pairs of spoken languages, in that the mappings between syntactic constructions from the source language to target language may not be one-to-one. In cases where the relationship is many-to-one (because the source language is underspecified in some way) the inherent ambiguity must be resolved to provide a grammatically acceptable translation.[1]

[1] An alternative form of the same problem occurs in the pronominal system where BSL distinguishes 'we-inclusive of hearer' from 'we-exclusive of the hearer'. For a translation of an English sentence containing 'we' the additional information must be inferred or volunteered by human intervention.

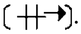
The sentence *"I put the bowls in the sink."* is synthesised to the following sign sequence.

```
[
[ [ bowl ],        [ 'b@Ul', non_raised ],
```

```
[ [ sink ],        [ 'sINk', non_raised ],
```

```
[ [ put ],         [ 'pUt', non_raised ],
```

```
[ [ dropped ],     [ ],                    [] ],
[ [ punct ],       [ ],                    [] ]
]
```

and the manner in which it is signed by the avatar is illustrated in Fig. 5. (Here, the agent pronoun 'I' has been omitted, as indicated by the empty sign 'dropped'.)

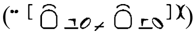
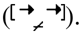
As noted above, the use of *placement* (allocation of positions in a 3D signing space around the signer to denote nominals) allows the opportunity for a signer to position objects and people at particular sign space locations. Subsequent pointing at these positions (*indexing*) is the equivalent of pronominal reference in spoken languages. However, groups of individuals and objects can be positioned at approximately the same position by a displaced repeated motion [3, 34]. Thus, nominals which can be pluralised in this way have a parameterised phonetic lexical entry, such as

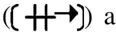
[bowl],['b@Ul', Brow],[, R]] ---> RHS

In the case of a singular nominal, R is instantiated to a HamNoSys empty list, but in the case of a plural nominal, R is instantiated to the repeated movement with a displacement .

Again as noted above, sign languages permit directional verbs to internally denote one or more of object, subject, and indirect object. However, in addition, they permit modification for agreement with a plural collective reading nominal by a sweeping motion prior to the movement (R1) or by repetition of the movement for a distributive reading (R2), as in the example below:

```
[[put],[ 'pUt', Brow ],[ ClassifierShape
        SourcePosition,
        R1,
        Motion,
        DestPosition,
        R2]   ] ---> RHS
```

Again, during sign language generation, the HPSG grammar constrains the generated form to be consistent with the allocation map positions for the elements, instantiating parameters to the appropriate HamNoSys subsequences. Hence, SourcePosition is taken from BOWLS ( ), DestPosition and orientation are taken from SINK ( ), classifier handshape is taken from BOWLS (), and the direction of motion is made consistent with the start and end locations ().

Moreover, the grammar requires agreement of a distributed nominal with instantiation of R2 to repeated displaced movement () and R1 to an empty HamNoSys sequence.[2] In the case that a collective interpretation is selected, this information is carried forward into the HPSG generation and R1 is instantiated to an appropriate sweep motion in HamNoSys and R2 to an empty HamNoSys sequence.

2.4 Plurals within the DRSs and HPSG semantic features

As discussed earlier, the synthesis of the HamNoSys for the phonetic sign language description begins from an HPSG semantic form that is consistent with the DRS generated from the original English sentence *"I put the mugs in the sink"*. During the DRS generation, the user volunteers information as to whether a plural is to be interpreted collectively or distributively. The generated DRS is a relatively standard formulation [14], as shown in Fig. 6.

The HPSG lexical entries and a small number of grammar rules, which enforce sign order and number agreement, constitute a collection of simultaneous constraints to which a valid sign sequence must conform.

3 Generating animation from HamNoSys/SiGML

3.1 Virtual human animation

A computer-generated avatar is modelled as a three-dimensional deformable surface mesh (Fig. 7 below). The surface mesh is a connected network of small coloured textured polygons; several thousands of these polygons

[2] On a technical note, the formulation of TAKE and PUT presented here originates from two different formulations for directional verbs. The description of the motion for TAKE is characterised by a HamNoSys 'replacement' of the location of the sign. Usually this form of replacement is used for describing the change in handshape within a sign rather than the location at which it is signed. Change of position is usually denoted by HamNoSys motion primitives aimed at a targeted location (destination) as with PUT here. The preliminary HPSG support to achieve this is now in place, but older lexical items need updating and testing to ensure this generalises to all forms of one- and two-handed motions is still to be undertaken. Essentially both formulations look to achieve the same effect, the former is simpler but exploits 'undocumented' features of HamNoSys, the latter is more complicated but more in the original spirit of HamNoSys.

Fig. 6 The DRS for *I put the mugs in the sink.* distributive interpretation

```
[ [e(1),v(0),v(3),v(4),a(2),a(3),a(4),q(0),c(0),t(1),set(0)]
  [ e(1) : put ( v(0) ,  v(4) ,  v(3) ) ,
    a(2) : me ( v(0) ) ,
    q(0) : every ( v(4) ) ,
    a(3) : bowl ( v(4) ) ,
    c(0) : plural ( v(4) ) ,
    t(1) : when ( e(1) ) ,
    t(1) = now ,
    set(0) = sumof ( v(4) ) ,
    a(4) : sink ( v(3) )
  ]
]
```

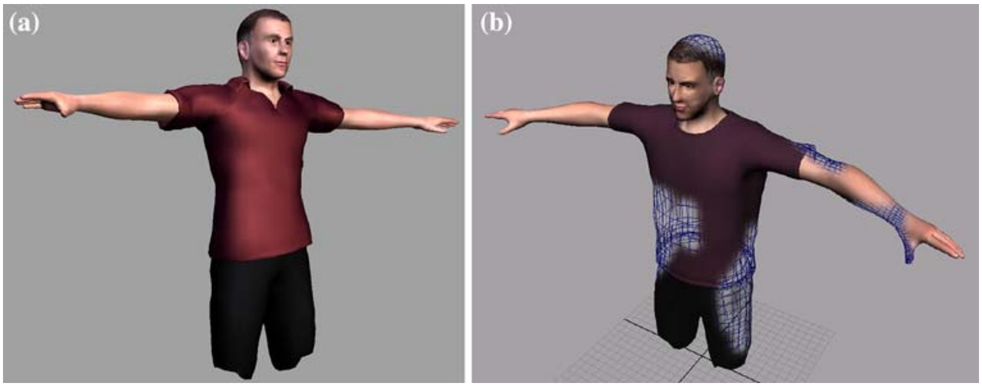


Fig. 7 The eSIGN avatar, VGuido

being needed to produce an acceptable level of realism. Given the coordinates of all the polygon vertices, this mesh can be displayed on the screen of a computer or mobile device using standard three-dimensional rendering techniques, which depend on a combination of graphics software and hardware. As well as its visible mesh, each avatar is endowed with a virtual skeleton, that is, a hierarchically structured set of virtual bones. The mesh is notionally attached to this skeleton, in the sense that the location and orientation of each polygon is defined with reference to those of one or more bones. This notional attachment means that any change in the (invisible) virtual skeleton configuration effectively determines a corresponding change in the configuration of the (visible) mesh. Hence, making a change to the skeleton configuration is an effective method of changing the avatar's posture as perceived on the display screen. The skeleton configuration can be specified with a very much smaller amount of data than the entire mesh, or a rendered two-dimensional image. Therefore, to transmit an animation over the net, the volume of data that must be transferred can be greatly reduced by transmitting only a sequence of skeleton configurations, and letting the end-user's machine generate the mesh and render it to the screen. To do this it needs to have the data specifying the attachment of the mesh to the skeleton, but this needs to be transmitted only once. Current consumer-level machines are able to construct and render the mesh in real time at a satisfactory frame rate.

An alternative method of changing the avatar's posture is through the application of *morphs*. A morph is a distortion of a region of the surface mesh, specified as a displacement of each vertex of the mesh within that region. This gives more detailed control over the shape of the mesh than is practical to achieve with bones, and is used for facial animation, in which subtle movements can have a large perceived effect on the expression.

In summary, given definitions of an avatar's skeleton, of its mesh, of the binding between the two, and of its facial morphs, rendering software and hardware can generate a real-time animation of that avatar from the appropriate stream of animation parameters, specifying for each animation frame the skeleton configuration and the weights for the facial morphs.

Avatar rendering technology has been developed for these projects by Televirtual Ltd, who produced the *Virtual Guido* (*VGuido*) avatar (Fig. 7) driven by their Mask-2 animation software. As well as having a high level of visual realism, the avatars provided by Televirtual for signing purposes have had to meet higher standards of anatomical realism than are required by most applications of avatar technology. This applies particularly in connection with the realism of the hands and fingers, for which signing is one of the most demanding applications.

The topology of the skeleton is largely compatible with the H-Anim standard, although *VGuido* uses a slightly more faithful representation of bones in the hand and does not need to animate legs and feet. The avatar definition also defines locations on the body and hands that may be placed in contact during signing. Although the technology has much in common with that used for human animation within MPEG-4, the main focus is on character-independent animation from high-level scripting notations, such as HamNoSys and SiGML, rather than from low-level animation parameters. During the ViSiCAST project it was shown that MPEG-4 Face and Body Animation could be used as an alternative to the bespoke avatar rendering technology.

3.2 *AnimGen*—synthesis of animation parameters

Although in linguistic terms HamNoSys or SiGML describe the required sign sequence at a relatively low level, this is nevertheless a significantly higher level than that of the stream of purely numeric animation parameters needed to drive an avatar as described earlier. The obligation to bridge this gap was one of the largest challenges implicit in the definition of the ViSiCAST text-to-sign-language pipeline. This challenge was met by a software module called *AnimGen*. AnimGen takes a (Gestural) SiGML sequence as input and produces a corresponding stream of low-level animation parameters as output. Even when restricted to the manual aspects of signing this is a large undertaking, salient aspects of which are described in [15, 16]. Among the more prominent issues which have to be addressed by this synthetic animation module are:

- Implementation of the large range of hand shapes which can be described in HamNoSys/SiGML.

These are currently implemented by a set of tables specifying joint angles for the 12 standard handshapes of HamNoSys, and rules for modifying them when HamNoSys hand shape modifiers are applied (specifying such things as the position of the thumb, or bendings of the fingers).

An inverse kinematic model of the hand would improve on this (since at present these lookup tables must be designed for each avatar separately), although for fast animation, such a model would be used when loading an avatar, to precompute the same tables.

- Establishing rules to translate what are often relatively loose specifications of absolute or relative positions of the hands into precise numerical coordinates—and ensuring the validity of these rules in all contexts.

Transcriptions originally written for people to interpret can be surprisingly hard to give a precise semantics to, even when human interpreters all agree on the meaning. This was a substantial part of the work involved in implementing HamNoSys.

- Identifying natural configurations and movements for parts of the signer's anatomy.

HamNoSys/SiGML is usually silent about positions of parts of the body such as arms and shoulders, because they do not function as phonetic articulators. By visually examining signing movements from live signers and videos, rules were designed to decide such things as how far the shoulder should move when reaching, or what direction the elbow should point. The rules uniquely determine all of the arm and torso joint rotations for a given positioning and orientation of the hands, by inverse kinematic algorithms.

- Generating natural motion, and avoiding collisions between body parts.

To obtain natural accelerations and decelerations when moving from one posture to another, a semi-abstract biocontrol model was used, whose parameters can be adjusted to produce the various *modalities* of HamNoSys: normal, sudden stop, tense, and lax. For each modality, this generates a table specifying, for any fraction of the elapsed time of a movement, what fraction of the spatial movement should be accomplished by that time.

Collision avoidance is simplified by the fact that it can be assumed that the postures being described by the input do not themselves involve penetration of body parts. It only needs to be ensured that the inverse kinematic and interpolation algorithms do not introduce any penetrations. This is achieved by a few more simple rules, such as one that requires the upper arm bone to always lie outside a certain cone, centred on the shoulder joint and approximating the shape of the upper torso on the same side of the body.

- Accomplishing all the above in real time.

The IK and collision-avoidance rules are designed to be simple enough to apply on every frame of the animation, while still giving realistic results.

The result is that when generating motion data for display at 25 frames per second, AnimGen requires only a very small fraction of the time budget.

AnimGen is also responsible for linking a sequence of signs together by interpolating from the final position of one sign to the initial position of the next. The final position of a sign must be held for a time, where required, and the motion linking signs is done using a modality that does not lead to the motion being considered as meaningful in itself.

The initial ViSiCAST prototype system dealt with a single avatar. As the system and its range of applications have developed, refinements have been made to the scripted animation system to make it as straightforward as possible to introduce new avatars, and to make it possible in principle for any avatar animation system to be driven by AnimGen, and hence to support scripted signing. These objectives have been met by identifying the relevant sets of avatar-dependent configuration data, and establishing formats for this configuration data, thus making the software itself essentially avatar-independent. This work has included the following aspects:

- Developing a flexible avatar-independent framework to support morph-based implementation of the facial non-manual features of HamNoSys 4.
- Establishing a standard for the definition of all those physical avatar characteristics which must be defined by an avatar supplier if that avatar is to be used for signing.
- Establishment of a standard format for avatar animation parameters.

Some aspects of this work are described in [5] and [6]. Significant effort has also been put into the ongoing task of improving the perceived quality of the signing, especially in the area of hand-shape production and in the configuration of the hands with respect to each other and to sites on the signer's body and locations in signing space. This effort has been driven partly by early user feedback, and especially by feedback from the sign language experts involved in the development of SiGML lexicons for demonstrator applications in the eSIGN project. The need for this effort can be regarded as a consequence of the decision to use the HamNoSys notation for the description of the sign language animations that are to be synthesised automatically by the software system. HamNoSys had previously been used only for sign language transcription, and therefore had been interpreted only by appropriately trained human readers.

In conclusion to this brief outline of the synthetic animation subsystem, it should be emphasised that work so far has been concerned solely with the generation of the phonetic-level aspects of signing, whether manual or non-manual. Although this would be very interesting, there has not yet been an opportunity to deal with prosodic aspects of sign language production, and still less with other kinds of expressive behaviour which may accompany a sign language performance and which, while not strictly linguistic, may nevertheless play an important communicative role.

3.3 Organisation of the animation subsystem

The sub-system forming the back-end of the ViSiCAST text-to-sign-language pipeline is based on three components: one concerned with the input of SiGML data and the translation of HamNoSys to SiGML, AnimGen, the synthetic animation module described in Sect. 3.2, and the avatar animation/rendering module described in Sect. 3.1. In addition, a control module is needed to coordinate these components, in particular, to cache the relatively large volumes of animation data generated by AnimGen and to schedule their timely transmission to the avatar animation module.

For the ViSiCAST and eSIGN projects, versions of these components were produced and packaged as ActiveX controls. Taken together these software modules effectively form a self-contained package which provides a relatively simple application programmer's interface (that is, API), which first allows the application software or website developer to place a panel containing a signing avatar within the window of a stand-alone application or within a web-page, and then allows sign language sequences expressed in SiGML to be performed on-demand by this signing avatar in real-time. This package has been used to build a stand-alone application providing a real-time HamNoSys/SiGML-to-signed-animation service accessible via a standard network connection to any other sofware module whether on the host computer system or on a remote system. This application acts as the back-end subsystem in the ViSiCAST text-to-sign-language pipeline. The package has also been used[3] by scripts in HTML pages to provide sign-language animation on the Web. More recently, a library of Java components with comparable capabilities and software interfaces has been developed, with a view to supporting a wider range of platforms.

4 Applications of sign language animation technology

4.1 Introduction

The development of the text-to-sign-language-animation system described in Sect. 2 achieves notable success with respect to certain syntactic phenomena over a relatively constrained lexicon. It delivered a prototype system which demonstrated the feasibility of taking spoken language input and generating sign language animation from this input. In doing this, significant ground had to be broken in a number of areas, notably in the HPSG-based sign-language modelling and generation, in defining the significant non-manual features of sign language, in the synthetic SiGML animation software, and in the development of realistic avatars. However, as with any prototype, the system allowed considerable scope for further development and improvement. Significant issues include the volume of

[3] (on computers running a Windows operating system, at any rate).

signed content, its quality, and the effort needed to generate it. The front-end of the system depends on the modelling of sign languages at an extremely precise and detailed level. The HPSG lexicons that this modelling produces are an extremely valuable resource. However, they take a relatively long time to develop: the sign language modelling which develops them therefore has to be regarded as a long-term undertaking.

This has provided motivation for the development of alternative methods for the creation of animated sign language material using the HamNoSys/SiGML scripted animation subsystem described in Sect. 3, especially for deployment on the Web. Some of this work has been undertaken as part of the Essential Sign Language Information on Government Networks (eSIGN) project [8]. This project was funded under the EU Framework V eCONTENT programme, its objective being the further development of ViSiCAST synthetic animation and language technology, together with supporting tools, in order to facilitate the effective deployment of virtual human signing by government information services, especially on the World Wide Web. Most of the partners in this project had previously played a significant part in the development of the ViSiCAST text-to-sign-language system.

The remaining section contains further details of the content creation process that supports the development of applications of sign language animation technology, followed by a brief account of some of the demonstrator applications developed in the eSIGN project.

4.2 Phonetic lexicon and signed content creation

In order for a Web page to display a signed animation performed by an avatar, the essential requirements are that the user has the avatar software installed on his/her computer, and that there is an SiGML description of the required sign language sequence. A small amount of quite routine scripting in an HTML page is then sufficient to make the avatar sign the given sequence, possibly in response to some user-generated trigger such as a click on a "Play" button. Thus the main effort required to generate new signed content relating to a given information domain is devoted to the creation of a lexicon[4] of SiGML sign definitions for the given domain. When undertaken by a person with the appropriate skill and experience, this need not be an unduly time-consuming task. On the other hand, some fairly unusual skills and experience are required: not only must the lexicon creator be familiar with the target sign language, but she/he must also know how to represent sign language at the phonetic level embodied in HamNoSys, and must be able to use the computer-based tools which allows this information to be input and recorded. Once lexicon entries have been created, they must also be validated by getting the avatar to sign them; of course, this step may require an entry to be re-worked if the result is unsatisfactory in some respect. The content-creating experience in the eSIGN project has been that, for any given individual with the appropriate background, the process of lexicon entry creation tends to be rather slow at first, but also tends to become much more efficient with increased experience.

A sample eSIGN lexicon entry is presented below. The SiGML text for the DGS "HAUS" sign introduced in Sect. 1.4 is shown in Fig. 8. The <sign_manual ...> element contains essentially the same information about the manual performance of the sign as the HamNoSys sequence in Fig. 1. Typically, this manual performance will be accompanied by mouthing derived from the German word "Haus": the appropriate sequence of visemes (that is, visible manifestations of spoken phonemes) is also specified in the SiGML lexicon entry, in the <sign_nonmanual> element.

Figure 9 shows a selection of the animation frames generated in the performance of the SiGML text in Fig. 8. The first frame shows the start of the sign, the second and third show how the depiction of the roof ends and that of the walls starts, and the fourth shows the downwards movement representing the walls. The middle pair of frames also show the mouth opening, as it performs the specified viseme sequence.

4.3 Sample applications

To give examples of the application of the developed synthetic animation technology, some of the demonstrators developed by members of the eSIGN project are briefly described. The DGS (German Sign Language) demonstration provides signed versions of material selected from http://www.hamburg.de, the official Web site of the city (state) of Hamburg. These signed pages can[5] be accessed at http://gebaerden.hamburg.de. The primary access point is a "Welcome" page (Fig. 10). This leads to material on the "Integrations-amt" (Integration Office), which deals with issues arising in relation to people with disabilities in their places of work; on Social Welfare legislation defining the rights of deaf people; on the Hamburg Lost Property Office;

[4] It should be emphasised that each entry in an eSIGN lexicon simply contains the phonetic information for a single fixed sign; an eSIGN lexicon is thus to be sharply distinguished from the HPSG lexicon described in Sect. 3, each of whose entries contains a much richer range of grammatical data about the sign language feature it describes.

[5] At the time of writing, 2006-07.

```
<sigml>
  <hamgestural_sign gloss="HAUS">
    <sign_nonmanual>
      <mouthing_tier>
        <mouth_picture picture="haUs"/>
      </mouthing_tier>
    </sign_nonmanual>
    <sign_manual both_hands="true" lr_symm="true">
      <handconfig handshape="flat"/>
      <handconfig extfidir="o"/>
      <handconfig palmor="dl"/>
      <handconstellation contact="touch"/>
      <directedmotion direction="dr"/>
      <tgt_motion>
        <changeposture/>
        <handconfig palmor="l"/>
      </tgt_motion>
      <directedmotion direction="d"/>
    </sign_manual>
  </hamgestural_sign>
</sigml>
```

Fig. 8 SiGML Notation for DGS sign "HAUS", including non-manual data

and on the Hamburg City Parliament, including an archive of press releases, speeches made by the president and vice-president, and details of sessions and votes in the Parliament.

As can be seen from the example in Fig. 10, as well as the signing avatar, VGuido, each page includes short excerpts of text. Using the "Play" buttons, the user can request to have the sign language (DGS) version of any of these pieces of text performed by the signing avatar. The user is thus able to choose which sequences to play, and in what order. There are also "Stop" and "Pause" buttons, so that the user is not forced to wait for longer sequences of signing to complete, once started. The user is also able to manipulate the virtual camera through which the avatar is viewed, changing both its position and its orientation relative to the avatar.

The aim of the eSIGN signed content creation for deaf people in the Netherlands was to assist them in improving their employment opportunities, and in gaining improved access to social services and other government-provided facilities. Hence, the signing avatar has been used to provide signed assistance to deaf people in completing several on-line forms, including forms needed in applying for interpreting services. Web pages have been developed to provide signed descriptions of job vacancies at the Viataal organisation, which provides services for deaf people in the Netherlands. A notable feature of these Web pages describing job vacancies is that they are created using the *Structured Content Creation Tool*. This is a tool developed to support the generation of Web pages which include signed content whose form is relatively tightly constrained, albeit in a way that often occurs quite naturally in the provision of information services. These constraints on the form of the signed content allow a SiGML lexicon to be created in advance. Rules are designed that enable correct sign language content to be generated from values selected via a form. Hence, the tool may be used by non-signers to generate Web pages which include signed content.

The eSIGN technology has also been used by Viataal to create "GebarenNet", a Web site for deaf people, shown in Fig. 11. This site showcases NGT (Dutch Sign Language) on the Internet using the eSIGN avatar, as part of a project sponsored by the national Dutch Ministry of Health, Welfare and Sport.

In the UK, signed content in BSL is provided via the Web site created by "Deaf Connexions", a voluntary-sector organisation providing support and services to deaf and hard of hearing people in the county of Norfolk in England. A button labelled with the eSIGN icon is displayed against those parts of the site for which a signed version is available. A click on this icon causes a separate window to appear, displaying the avatar signing the given section.

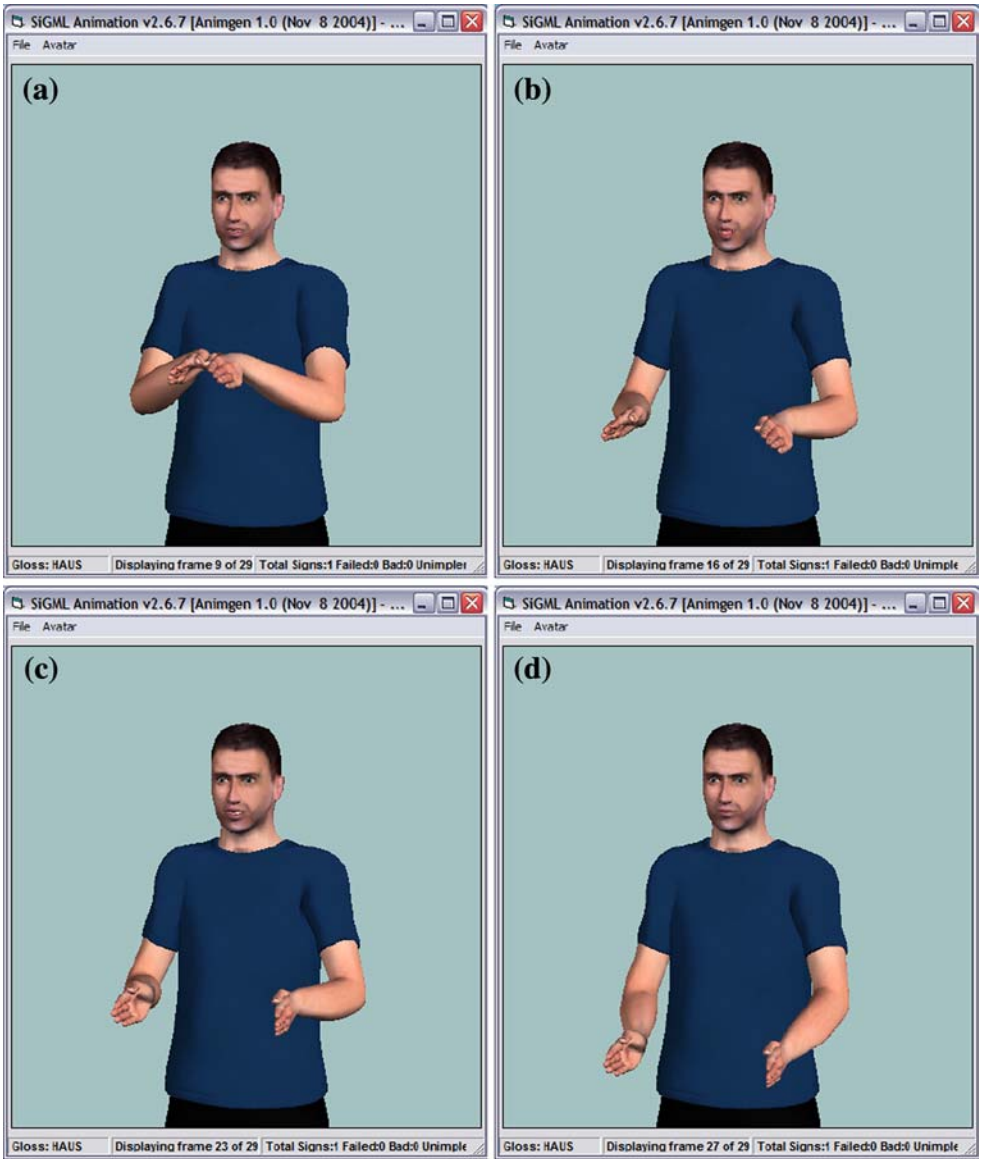
Another application of synthetic signing technology in the UK is the Voice Activated Network Enabled Speech and Sign Assistant (VANESSA) system [11]. This system uses the signing avatar to augment communication with deaf clients in a Council Information Centre in the Forum in the centre of the city of Norwich. Here, the signing avatar provides assistance in filling in paper forms such as those for Housing Benefit, in arranging a booking for an interpreter, and in some commonly occurring situations such as redirecting a client to an office appropriate to his or her query.

Some information on user evaluation of these applications appears in [17].

5 Conclusions

Progress on virtual human signing continues to reflect the multi-disciplinary nature of the application area, and the theoretical and practical advances in graphical portrayal of virtual humans and language processing in a visual modality. The work described in the previous sections has shown

- that it is possible to use suitably constructed avatars to produce intelligible sign-language performance in real time, driven by phonetic-level descriptions expressed either in HamNoSys or Gestural SiGML;
- that high-level linguistic analysis and HPSG-based sign language modelling techniques support semi-automatic generation of high-quality translations of English (or other spoken-language) text into sign languages such as BSL, producing phonetic-level representations which can be performed by a signing avatar;
- that the phonetic-level interface to the signing avatar also supports the creation at a relatively low cost of sign language content for performance on Web sites, and in other applications, through the use of less grammatically sophisticated sign lexicons.



Gloss: HAUS | Displaying frame 9 of 29 | Total Signs:1 Failed:0 Bad:0 Unimplei

Gloss: HAUS | Displaying frame 16 of 29 | Total Signs:1 Failed:0 Bad:0 Unimple

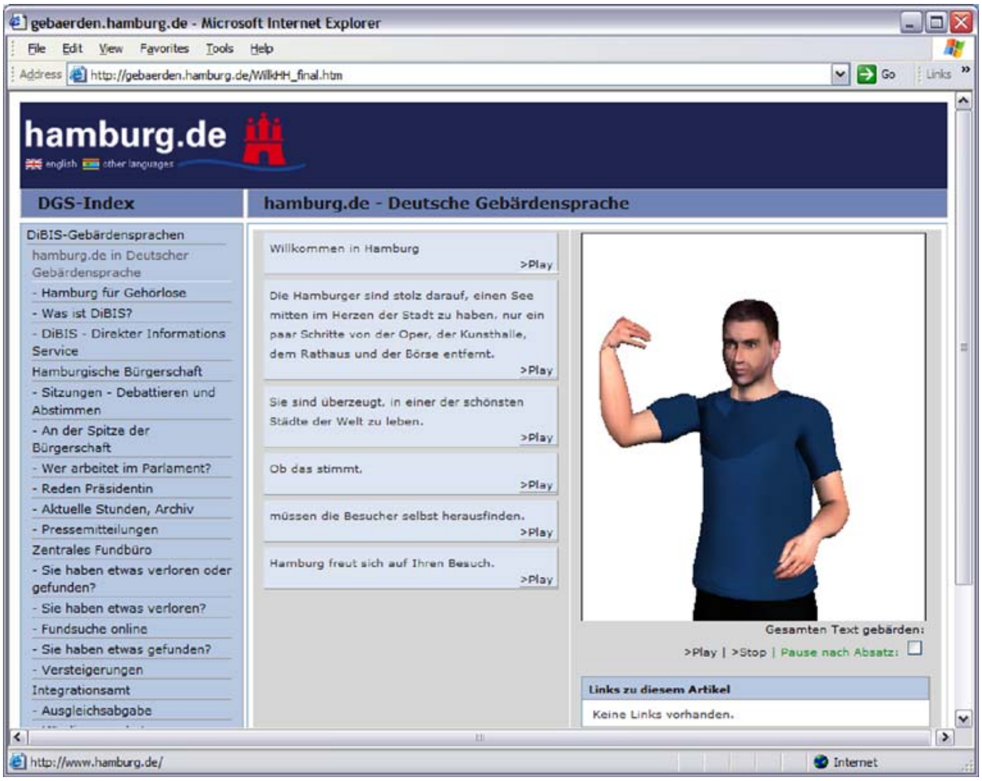
Gloss: HAUS | Displaying frame 23 of 29 | Total Signs:1 Failed:0 Bad:0 Unimple

Gloss: HAUS | Displaying frame 27 of 29 | Total Signs:1 Failed:0 Bad:0 Unimple

Fig. 9 Snapshots of VGuido's performance of the SiGML text for the DGS sign "HAUS"

The detail of current research characterisations of sign language user behaviour and the sign language grammar informed by such research are undoubtedly crude, requiring significant programs of work for further refinement and verification. However, it is clear that simplistic transliteration between a spoken and a sign language using Sign Supported variants produces unnatural signing comparable to unnatural word-by-word translations between spoken languages. Hence, (semi-) automated sign language generation requires serious studies of sign language constructs at sufficient levels of detail to allow the kinds of formulation discussed above to be undertaken. Synthesis of HamNoSys descriptions within an HPSG framework has been productive in exploring the sign language generation side of this problem. Most specifically, formulations of HamNoSys for directional verbs have to be in terms of specific styles of HamNoSys descriptions (illustrated above) in order to appropriately parameterise them with the necessary inherited information from other linguistic constructs. The HPSG formalism assists greatly in this.

Visualisation of synthetically generated sign language by means of a signing avatar potentially supports a methodology whereby hypotheses concerning sign language grammar can be reviewed by native signers and revised in the light of their feedback. The exploitation of this methodology is still at an early stage, and the shortcomings of these early models are readily apparent. However, linguistic tradition has progressed by virtue of revisions of


gebaerden.hamburg.de - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://gebaerden.hamburg.de/WilkHH_final.htm
Go
Links
hamburg.de
english other languages
DGS-Index
hamburg.de - Deutsche Gebärdensprache
DiBIS-Gebärdensprachen
hamburg.de in Deutscher Gebärdensprache
- Hamburg für Gehörlose
- Was ist DiBIS?
- DiBIS - Direkter Informations Service
Hamburgische Bürgerschaft
- Sitzungen - Debattieren und Abstimmen
- An der Spitze der Bürgerschaft
- Wer arbeitet im Parlament?
- Reden Präsidentin
- Aktuelle Stunden, Archiv
- Pressemitteilungen
Zentrales Fundbüro
- Sie haben etwas verloren oder gefunden?
- Sie haben etwas verloren?
- Fundsuche online
- Sie haben etwas gefunden?
- Versteigerungen
Integrationsamt
- Ausgleichsabgabe
>Play
Gesamten Text gebärden:
>Play | >Stop | Pause nach Absatz: □
Links zu diesem Artikel
http://www.hamburg.de/
Internet

Willkommen in Hamburg

Die Hamburger sind stolz darauf, einen See mitten im Herzen der Stadt zu haben. nur ein paar Schritte von der Oper, der Kunsthalle, dem Rathaus und der Börse entfernt.

Sie sind überzeugt. in einer der schönsten Städte der Welt zu leben.

Ob das stimmt.

müssen die Besucher selbst herausfinden.

Hamburg freut sich auf Ihren Besuch.

Keine Links vorhanden.

Fig. 10 http://gebaerden.hamburg.de WWW site—“Welcome”

theories to account for new and more extensive data. The advent of virtual signing provides the opportunity for formulations of sign language grammars to become more open to comment and correction by native signers who need not be fluent in linguistic theory.

It is clear that translation between spoken and sign language faces no less problems than translation between spoken languages. Automatic translation of the latter as yet yields only modest quality. In addition, the most widely used translation techniques are based on large language-specific corpora, and there are still no sign language corpora of comparable size. In the immediate future, high-quality creation of large volumes of signed information will remain a labour-intensive task.

The use of the HamNoSys notation and its SiGML realisation as the phonetic-level interface between the major components of signing systems demonstrates the practical application of a notation originally designed for linguistic analysis to the task of synthetic generation. The bandwidth efficiency of such notations for storage and transmission by comparison with unanalysed video data opens up the most promising prospect for the collection of large corpora of signed data from which future linguistic analysis may proceed.

Although the authors believe that the system reported in this paper is the most advanced attempt to translate aspects of written language to sign, a number of other systems provide animation of sign language using virtual humans by synthesis from notation, sometimes in combination with hand-crafted elements.

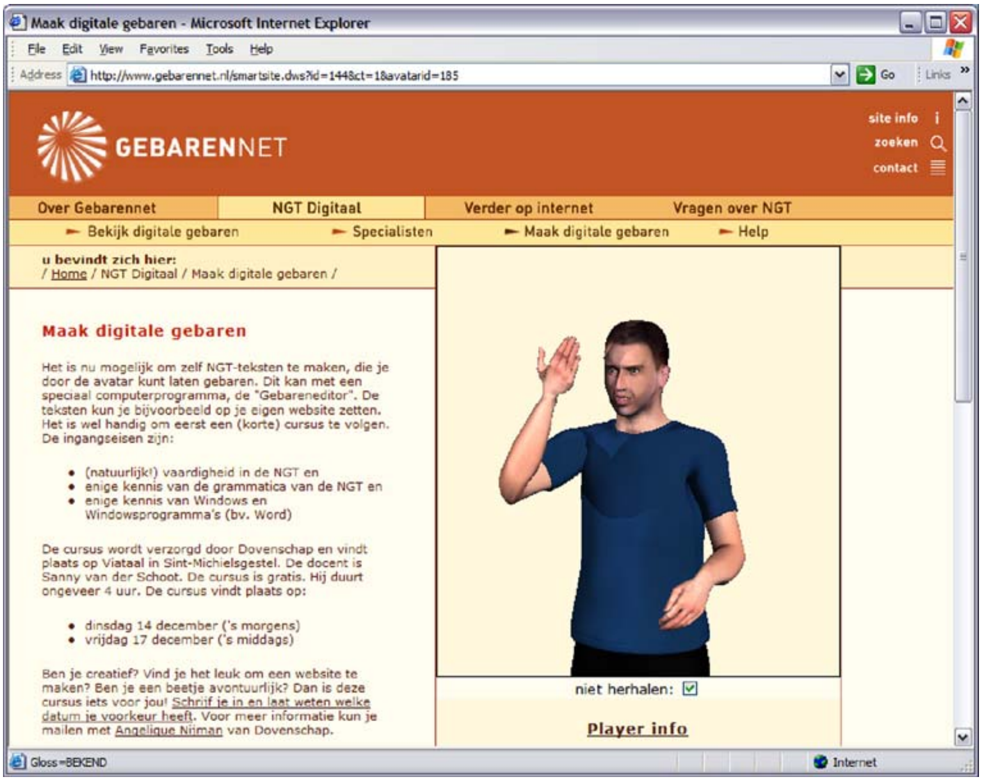
Signing-Avatar and *Sign Smith Studio* from VCom3D[6], may be used for performing American Sign Language (ASL) on web pages. It has some similarities with the approach here: synthesis is based on sign representation using the Liddell/Johnson system, encoded in XML, but can also include motion capture data or keyframe animation, especially for complex signs which cannot be described easily in the notation system. Animation is performed in VRML[7]. A system for automatic translation from English to signing is provided that targets Signed English rather than ASL.

SignSynth [12] is a signing animation system based on the Stokoe notation, and also addressing the issue of translating English text to sign language.

Systems concerned with animation and translation for other natural sign languages include the GesSyCa system [18] for French Sign Language, which uses the QualGest

[6] http://www.vcom3d.com

[7] Virtual Reality Modelling Language, the open standard (now succeeded by X3D) for describing 3D interactive animated worlds. See http://www.web3d.org.



u bevindt zich hier:
/ Home / NGT Digitaal / Maak digitale gebaren /

Maak digitale gebaren

Het is nu mogelijk om zelf NGT-teksten te maken, die je door de avatar kunt laten gebaren. Dit kan met een speciaal computerprogramma, de "Gebareneditor". De teksten kun je bijvoorbeeld op je eigen website zetten. Het is wel handig om eerst een (korte) cursus te volgen. De ingangseisen zijn:

- (natuurlijk!) vaardigheid in de NGT en
- enige kennis van de grammatica van de NGT en
- enige kennis van Windows en Windowsprogramma's (bv. Word)

De cursus wordt verzorgd door Dovenschap en vindt plaats op Viataal in Sint-Michielsgestel. De docent is Sanny van der Schoot. De cursus is gratis. Hij duurt ongeveer 4 uur. De cursus vindt plaats op:

- dinsdag 14 december ('s morgens)
- vrijdag 17 december ('s middags)

Ben je creatief? Vind je het leuk om een website te maken? Ben je een beetje avontuurlijk? Dan is deze cursus iets voor jou! Schrijf je in en laat weten welke datum je voorkeur heeft. Voor meer informatie kun je mailen met Angelique Nijman van Dovenschap.

Fig. 11 WWW site for deaf people in the Netherlands (http://www.gebarennet.nl)

notation system, and THETOS [9, 33] for animating Polish Sign Language and translating Polish into PSL.

Virtual human developments continue apace in the computer games arena, and these developments will continue to raise the quality of virtual humans in general. However, such applications do not focus upon hand shape and hand configurations, nor facial expression to the extent needed for high-quality virtual signing. The work on virtual human signing technology reported in this paper has produced the most completely developed architectural framework and algorithms to date in support of real-time sign language applications. It is important that research continues to further develop the anatomical models and realistic synthetic motion algorithms which can give both greater realism and greater coverage of communicatively significant features.

Acknowledgments The authors acknowledge with gratitude, funding from the European Union for their work on the ViSiCAST and eSIGN projects under the European Union Framework V programme; we are also grateful for assistance from our partners in these projects, including Televirtual Ltd., who developed the VGuido avatar for the eSIGN project together with the accompanying avatar rendering software.

References

1. Blackburn, P., Bos, J.: Representation and inference for natural language. In: A First Course in Computational Semantics, vol II. http://www.coli.uni-sb.de/~bos/comsem/book1.html (1999)
2. Bray, T., Paoli, J., Sperberg, C.M., Mahler, E., Yergeau, F. (eds.): Extensible Markup Language (XML) 1.0, 3rd edn. http://www.w3.org/TR/REC-xml/. Retrieved 7 July 2006 (2004)
3. Brien D. (Ed.): Dictionary of British Sign Language/English. Faber and Faber, London, Boston (1992)
4. Carpenter, B., Penn, G.: The Attribute Logic Engine. User's Guide. Version 3.2 Beta, Bell Labs (1999)
5. Elliott, R., Glauert, J.R.W., Jennings, V.J., Kennaway, J.R.: An overview of the SiGML notation and SiGMLSigning software system. In: Streiter, O., Vettori, C. (eds.) Workshop on Representing and Processing of Sign Languages, LREC 2004, Lisbon, pp 98–104 (2004)
6. Elliott, R., Glauert, J.R.W., Kennaway, J.R.: A framework for non-manual gestures in a synthetic signing system. In: Keates, S., Clarkson, P.J., Langdon, P., Robinson, P., (eds.) 2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Cambridge, 2004, pp. 127–136 (2004)
7. Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I.: Development of language processing support for the ViSiCAST Project. In: ASSETS 2000, 4th International ACM SIGCAPH Conference on Assistive Technologies, Washington DC (2000)
8. eSIGN Project (2004). Project WWW site at http://www.sign-lang.uni-hamburg.de/eSIGN/
9. Francik, T., Fabian P.: Animating sign language in the real time. In: 20th IASTED International Multi-Conference on Applied Informatics, 2002, pp. 276–281 (2002)
10. Glauert, J.R.W.: ViSiCAST: Sign language using virtual humans. In: International Conference on Assistive Technology (ICAT 2002), Derby, pp. 21–33 (2002)
11. Glauert, J.R.W., Elliott, R., Cox, S.J., Tryggvason, J.T., Sheard, M.: VANESSA—A System for Communication between Deaf and Hearing People. Technology and Disability (Special issue on Virtual Reality and Disability) (2006, in press)

12. Grieve-Smith, A.B.: SignSynth: A sign language synthesis application using Web3D and Perl. In: Wachsmuth, I., Sowa, T. (eds.) 4th International Workshop on Gesture and Sign Language Based Human–Computer Interaction, LNAI 2298. Springer, Heidelberg, pp. 134–145 (2001)
13. Hanke, T.: HamNoSys—representing sign language data in language resources and language processing contexts. In: Streiter, O., Vettori, C. (eds.) Fourth International Conference on Language Resources and Evaluation (LREC 2004). Representation and Processing of Sign Languages Workshop, pp. 1–6. European Language Resources Association, Paris (2004)
14. Kamp, H., Reyle, U.: From Discourse to Logic. Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer, Dordrecht (1993)
15. Kennaway, J.R.: Synthetic animation of deaf signing gestures. In: Wachsmuth, I., Sowa, T. (eds.) 4th International Workshop on Gesture and Sign Language Based Human–Computer Interaction, LNAI 2298, pp. 146–157. Springer, Heidelberg (2001)
16. Kennaway, J.R.: Experience with and requirements for a gesture description language for synthetic animation. In: Camurri, A., Volpe, G. (eds.) Gesture-based Communication in Human–Computer Interaction, LNAI 2915. Springer, Heidelberg (2004)
17. Kennaway, J.R., Glauert, J.R.W., Zwitserlood, I.: Providing signed content on the Internet by synthesized animation. ACM Trans Comput Hum Interact (TOCHI) (2007, in press)
18. Lebourque, T., Gibet, S.: A complete system for the specification and the generation of sign language gestures. In: Braffort, A., Gherbu, R., Gibet, S., Richardson, J., Teil, D. (eds.) 3rd International Workshop on Gesture-Based Communication in Human-Computer Interaction, LNAI 1739, pp. 227–238. Springer, Heidelberg (1999)
19. Liddel, S.K.: Structures for representing handshape and local movement at the phonemic level. In: Fischer, S.D., Siple, P. (eds.) Theoretical Issues in Sign Language Research, vol. 1, pp. 37–65. University of Chicago Press, Chicago (1990)
20. Marshall, I., Safar, E.: Extraction of semantic representations from syntactic CMU link grammar linkages. In: Angelova, G. (ed.) Recent Advances in Natural Language Processing (RAN-LP), pp. 154–159, Tzigov Chark. ISBN 954-90906-1-2 (2001)
21. Marshall, I., Safar, E.: Sign language synthesis using HPSG. In: Ninth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI), Keihanna (2002)
22. Marshall, I., Safar, E.: Sign language generation in an ALE HPSG. In: Stefan, M. (ed.) Proceedings of the HPSG04 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven, pp. 189–201. CSLI Publications. http://csli-publications.stanford.edu/ (2004)
23. Neidle, C., Kegl, J., MacLaughlin, D, Bahan, B., Lee, R.G.: The Syntax of American Sign Language. MIT Press, Cambridge (2000)
24. Pezeshkpour, F., Marshall, I., Elliott, R., Bangham, J.A.: Development of a legible deaf signing virtual human. In: Proceedings IEEE International Conference on Multimedia Computing and Systems, Florence, June 1999, vol 1 (1999)
25. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J., et al. Hamburg Notation System for Sign Languages—An Introductory Guide. International Studies on Sign Language and the Communication of the Deaf (5). Institute of German Sign Language and Communication of the Deaf, University of Hamburg, Hamburg (1989)
26. Pollard, C., Sag, I.A.: Head-Driven Phrase Structure Grammar. The University of Chicago Press, Chicago (1994)
27. Safar, E., Marshall, I.: The architecture of an english-text-to-sign-languages translation system. In: Angelova, G. (ed.) Recent Advances in Natural Language Processing (RANLP), Tzigov Chark, 2001, pp. 223–228 (2001a)
28. Safar, E., Marshall, I.: Translation of english text to a DRS-based sign language oriented semantic representation. In: Conference sur le Traitment Automatique des Langues Naturelles (TALN), vol. 2, pp. 297–306 (2001b)
29. Safar, E., Marshall, I.: An intermediate semantic representation extracted from english text for sign language generation. In: Seventh Symposium on Logic and Language, Pecs (2002)
30. Safar E., Marshall I.: Translation via DRT and HPSG. In: Gelbukh, A. (eds.) Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), pp. 58–68. LNCS. Springer, Mexico City (2002)
31. Sleator, D., Temperley, D.: Parsing English with a Link Grammar. Carnegie Mellon University Computer Science Technical Report CMU-CS-91-196 (1991)
32. Stokoe, W.C.: Sign Language Structure, 2nd edn. Linstok Press, Silver Spring (1978)
33. Suszczanska, N., Szmal, P., Francik, J.: Translating Polish texts into sign language in the TGT system. In: 20th IASTED International Multi-Conference on Applied Informatics, 2002, pp. 282–287 (2002)
34. Sutton-Spence, R., Woll, B.: The linguistics of British sign language. An Introduction. University Press, Cambridge (1999)