The Development of Language Processing Support for the ViSiCAST Project

Ralph Elliott, John Glauert, Richard Kennaway, Ian Marshall
[+ Kevin Parsons, Éva Sáfár]

{re,jrwg,jrk,im}@sys.uea.ac.uk

School of Information Systems,
UEA Norwich, UK

ASSETS 2000, Arlington VA, 2000-11-14

UEA NORWICH

Outline

- *ViSiCAST – Introduction/Background*
- *Language Processing in ViSiCAST*
  - General Approach
  - Natural Language to Semantics
  - Signing Gesture Language

ViSiCAST Project

- Virtual Signing – Capture, Animation, Storage and Transmission
- Aim: "…support improved access by deaf citizens to information and services in sign language".
- Funded under EU Framework V Programme [+ ITC and PO in UK]
  - "pre-competitive" research

ViSiCAST – Background

- *Builds on Two Earlier UK Projects…*
- *(ITC) Simon-the-Signer (97-99)*
  - ITC (UK Independent Television Commission), Televirtual, UEA Norwich
- *(PO) Tessa (98-00)*
  - Post Office, Televirtual, UEA Norwich
- *Both based on virtual human signing*
  - using Televirtual's motion-capture driven avatar technology

Motion-Capture Based Virtual Human Signing

- *Motion Capture Streams*
  - body
    - *magnetic tracking*
  - face
    - *reflective markers + head-mounted camera*
  - hands
    - *gloves with bend-sensors*

Virtual Humans (Avatars)

- *Bones-Set*
  - lengths
  - interconnection topology ("joints")
  - configure: by specifying angle and orientation at each joint
- *Rendering*
  - attach mesh ("wire-frame") to Bones-set
  - apply texture-mapping to mesh
- *Animation*
  - sequence of rendered frames
  - each defined by a Bones-Set configuration

From Capture to Signing (Simon & Tessa)

- *Capture "clips" of signing*
  - based on vocabulary for chosen subject area
  - requires calibration – match signer to avatar
- *Segment/Edit clips*
  - save as files, one per sign
- *Generate Stream of Sign Names*
  - for required script
- *Feed Sign Stream to Avatar*
  - acts as a "Player" for Stream (with blending)

"Sign Supported" vs. "Authentic" Sign Languages

- *In UK:*
  - SSE Sign-Supported English
    - *one sign per word (approx.)*
    - *follow English word order*
  - BSL British Sign Language
    - *one sign per concept*
    - *use of "signing space" around signer's body*
    - *has own word order, morphology*
  - SSE and BSL both utilize finger-spelling

Simon & Tessa

- *Simon-the-Signer* [Broadcast TV]
  - generate signed accompaniment to broadcast, using Teletext stream as source
  - SSE
- *Tessa* [Retail, PO]
  - convert counter-clerk's voice input to text, using speech recognizer
  - generate sign stream from text
  - BSL, but limited repertoire

ViSiCAST Partners (UK)

- *ITC*
- *Post Office*
- *Televirtual, Norwich*
- *School of Information Systems, Norwich*
- *RNID*
  - Royal National Institute for Deaf People

ViSiCAST Partners - contd.

- *IDGS, University of Hamburg*
  - Institute for German Sign Language and Communication of the Deaf
- *IRT, München*
  - Institute für Rundfunk Technik
- *INT, Evry (Paris)*
  - Institute National des Télécommunications
- *IvD, Sinkt-Michelsgestel (Netherlands)*
  - Instituut voor Doven

ViSiCAST: Application Areas

- Broadcasting
- Retail - " face-to-face"
- WWW

ViSiCAST: Development of Supporting Technologies

- *Avatar Technology*
- *Language Processing*

NL Processing – ViSiCAST Approach

- *Develop* semi-automated *translation system*
  - automated transformations
  - augmented by user-interaction …
    - *to resolve ambiguity*
      - e.g. “ give” , “ inject”
    - *to improve quality*

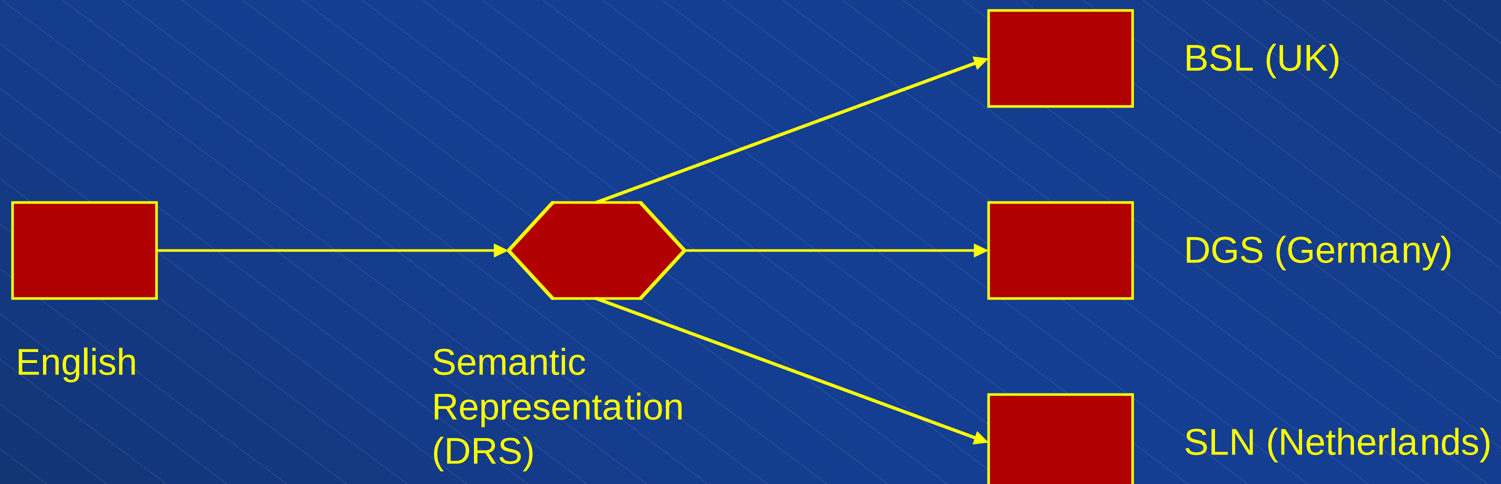
Stages on Path from NL to Signing

1. NL (English)
2. Semantic Representation
3. Morphology (Sign-Language Specific)
4. Signing Gesture Notation (SiGML)
5. Animation

… Compare/Contrast with pre-ViSiCAST:

- *Off-line preparation*
  - Motion Captured clips of signing
  - Segmentation/Editing of clips
- *From Script to Signing*
  - From Text to Stream of Sign File Names
  - Feed Sign Stream to Avatar as “ Player”

ViSiCAST: Route To National Sign Languages



Stages: NL to Semantic Representation

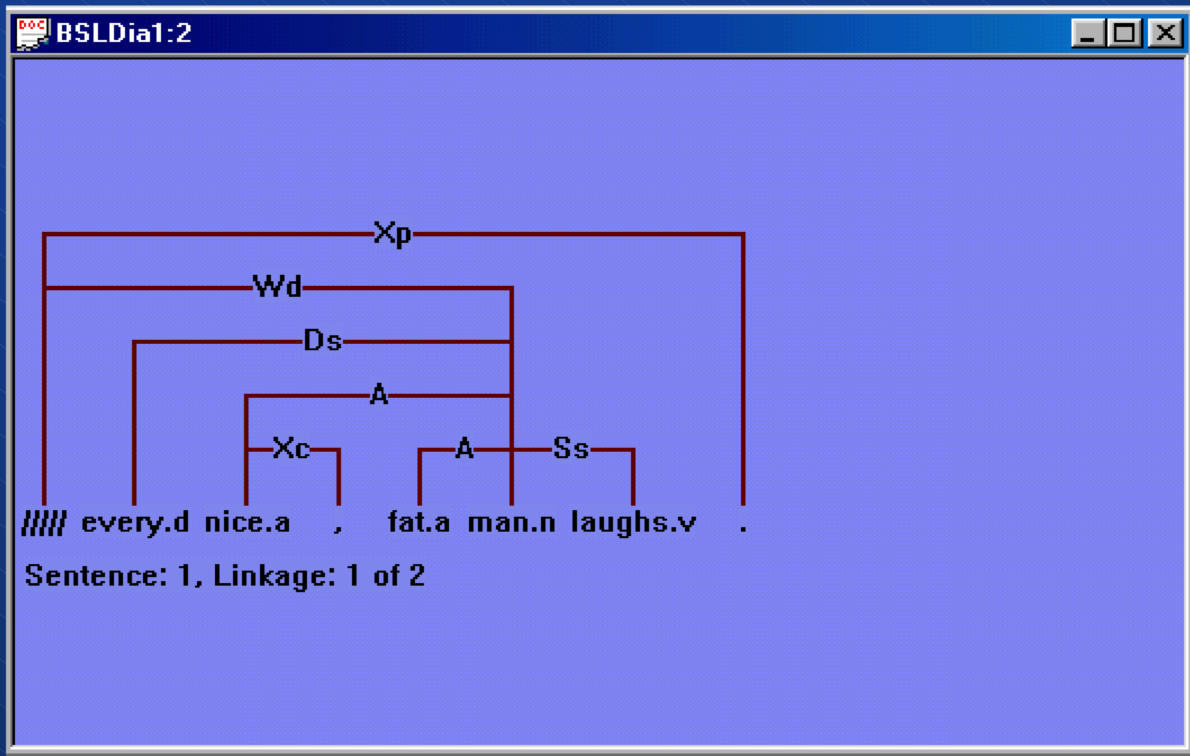
1. NL (English)
2. Semantic Representation
3. Morphology (Sign-Language Specific)
4. Signing Gesture Notation (SiGML)
5. Animation

Natural Language Parsing

- *Use "Link Grammars" Parser*
  - Sleator & Temperley, CMU
- *Represent each sentence as a linkage:*
  - a set of links
- *Each link:*
  - identifies a specific grammatical relationship between a pair of word occurrences in the sentence

CMU Linkage Diagram

- *" Every nice, fat man laughs."*



Linkage as a Set of 7-tuples

- *[[{m, 5, 0, Wd, Wd, Wd, 5},*
  {{}, 10, 0, Xp, Xp, Xp, 10},
  {m, 4, 1, Ds, Ds, Ds, 5},
  {m, 1, 2, Xc, Xc, Xc, 3},
  {m, 3, 2, A, A, A, 5},
  {m, 1, 4, A, A, A, 5},
  {m, 1, 5, Ss, Ss, Ss, 6},
  {m, 1, 6, MV, MVp, MVp, 7},
  {m, 2, 7, J, Js, Js, 9},
  {m, 1, 8, Ds, Ds, Ds, 9},
  {{}, 1, 10, RW, RW, RW, 11}]]

Semantic Representation

- *Based on Discourse Representation Theory (DRT) [Kamp & Reyle, 1993]*
- *Represent sentences:*
  - modified form of Discourse Representation Structures [DRSs]
  - "nested-box" representation ...

Box Representation for DRS

- U: *set of referents (variables) presently in use*
- Con: *set of conditions constraining the referents*

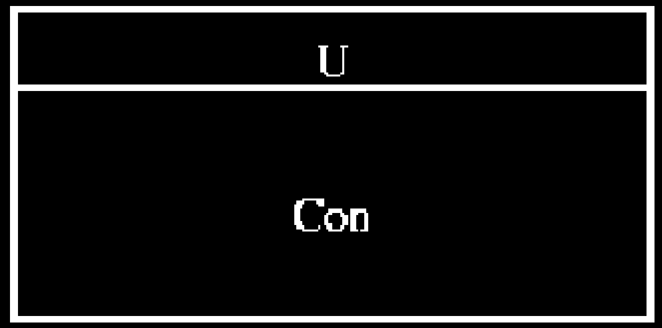


Fig 1 The template for a simple DRS.

Features of DRS Scheme

- *Each proposition is labelled*
  - allows incorporation of temporal information:
    - *t1: when(e1), t1=now, e1: happy(Mary)*
- *Use λ-terms to represent DRS fragments with place holders*
- *Supports distinctive characteristics of SLs:*
  - "Topic-Comment" structure
  - "Directional" verbs
    - *e.g. "give" (who-whom?)*

Route from NL Sentence to DRS

- Sentence $\rightarrow$ CMU Parser Linkage
- Place links in order for construction
- Look up λ-abstraction for each link
- Reduce (β-convert and DRS-merge) to obtain final DRS

Transformation to DRS – Example

- *" Every nice man laughed."*
- *Links for "* every nice man*" :*

    [m, 1, 2, A, A, A, 3] nice-man

    [m, 2, 1, Ds, Ds, Ds, 3] every-man

    [m, 3, 0, Wd, Wd, Wd, 3] ////-man

... in order of processing

λ-Term Example

- *λ-term corresponding to adjective "nice":*
  - lambda(P, *//property*
    lambda(Y, *//referent*
    merge(drs([], [Lab:Cond]), P@Y)
)
) *where* Cond=*nice*(Y)

(a) Apply Noun to Adjective

```
• lambda(_G14416,          // Y
      merge(
       drs([],
        [attr(_G14414): nice(_G14416)]),
       drs([],
        [a(_G14598):man(_G14416)])
      )
   )
```

(b) Apply Result (a) to Determiner

```
• lambda(_G14509,        // verb phrase
      drs([],
       [merge(
        drs([v(_G14504)],        // v0
         [q(_G14502):forall(v(_G14504))]),
        merge(
         drs([],
          [attr(_G14414):nice(v(_G14504))]),
         drs([], [a(_G14598):man(v(_G14504))])
  ))
  > (_G14509@v(_G14504))]))
```

(c) Apply Verb to Result (b)

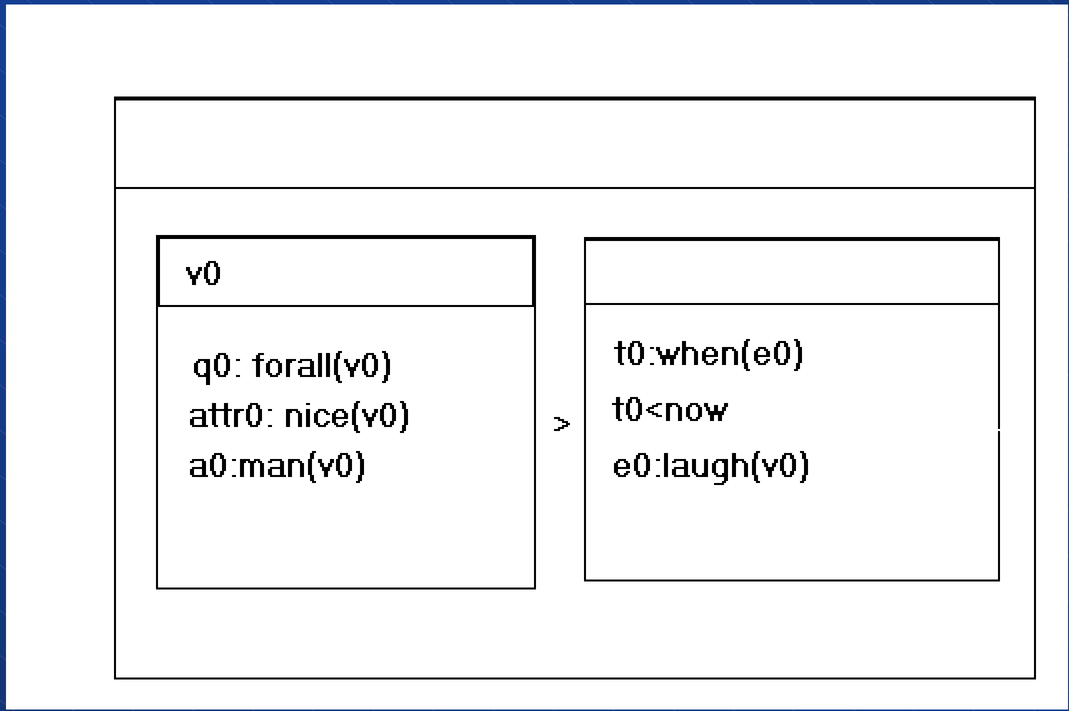
```
• drs([],
      [merge(
        drs([v(_G14504)],
         [q(_G14502):forall(v(_G14504))]),
        merge(
         drs([], [attr(_G14414):nice(v(_G14504))]),
         drs([], [a(_G14598):man(v(_G14504))])))
      >drs([],
         [t(_G17334):when(e(_G17332)),
         t(_G17334)<now, e(_G17332):laugh(v(_G14504))]
        )
      ]
  )
```

Final DRS for Example

- *“ Every nice man laughed.”*

- ```
drs([], [drs([v(0)],
              [q(0):forall(v(0)), attr(0):nice(v(0)),
               a(0):man(v(0))])
          >drs([],
              [t(0):when(e(0)), t(0)<now, e(0):laugh(v(0))]
)
        ]
)
```
```

Box Diagram for Final DRS in Example



Current Status – Coverage

- Transitive/intransitive verbs
- Temporal auxiliaries
- Passive verbs
- Imperative sentences
- Prepositional phrases on nouns and verbs (location only)
- Adjectives (any number)
- Determiners (indefinite, definite)
- Pronouns (but work on co-reference is in progress)
- Relative clauses (subject and object)
- Questions
- Proper Nouns

Stages – Morphology

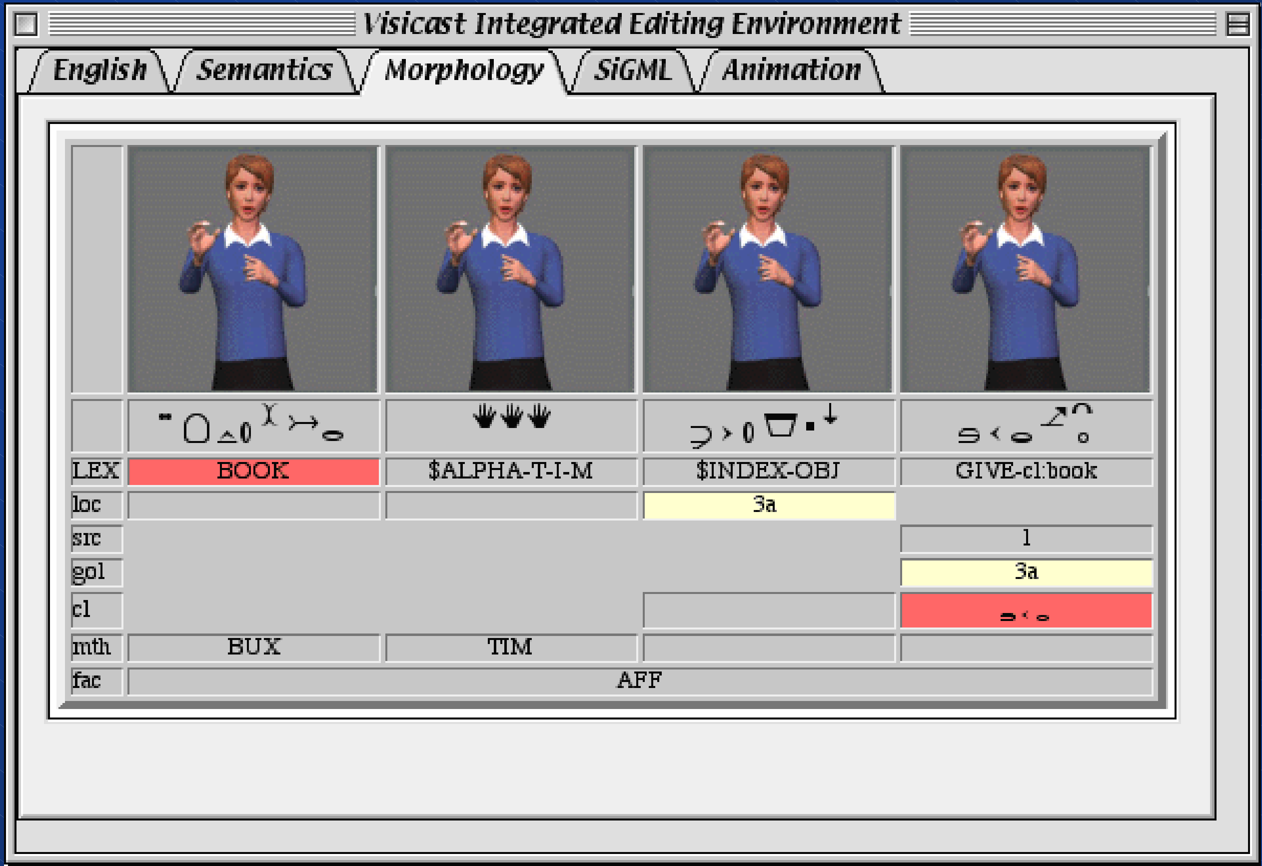
1. *NL (English)*
2. *Semantic Representation*
3. *Morphology (Sign-Language Specific)*
4. *Signing Gesture Notation (SiGML)*
5. *Animation*

- *e.g. Morphology for:*

  "Indeed, I'll give the book to Tim."

Morphology – (Projected) Representation

[Example due to Thomas Hanke, IDGS, U Hamburg]



Stages – SiGML

1. NL (English)
2. Semantic Representation
3. Morphology (Sign-Language Specific)
4. Signing Gesture Notation (SiGML)
5. Animation

SiGML

- *Signing Gesture Markup Language*
- *Based on:*
  - HamNoSys – Hamburg Notation System
  - XML – Extensible Markup Language

HamNoSys

- *General notation for signing*
  - originally defined primarily for purposes of recording, transcription, study of signing
- *Intention:*
  - capable of representing any sign language
    - *but a few enhancements in area of non-manual features are needed*
- *Defines*
  - semantic model for signing gestures
  - "pictographic" notation

HamNoSys Semantic Model – Manual Gestures

- *Hand Configuration*
- *Location*
  - in “signing space”
  - i.e. relative to the body of the signer
- *Motion*
  - i.e. “actions” of various kinds
    - *change configuration and/or location*

Hand Configuration

- *Hand Shape – hundreds of them*
- *Hand Orientation*
  - “ finger base orientation”
  - “ palm orientation”

Location (i)

- *Positions on head and body*
  - e.g. top of head, nose, neck, chest level etc.
- *Modifiers indicate*
  - position on "left-centre-right" spectrum
  - "contact distance"
    - *touching, close, normal, far*

Location (ii)

- *Positions on (non-dominant) arm and hand*
  - e.g. upper arm inside of elbow, ball of thumb, middle-joint-of-ring finger

Motion – Main Features

- *Absolute – i.e. "targeted"*
  - new hand position and/or
  - new hand configuration
- *Relative*
  - direction of motion from initial configuration
  - implicit target
    - *... a "normal" distance*

Motions – Composition

- *Temporal Sequence*
  - of distinct motions and/or
  - repetition of a single motion
    - *single or multiple*
- *Parallel*
  - i.e. several motions over a single temporal interval

Directed Motion – Variants

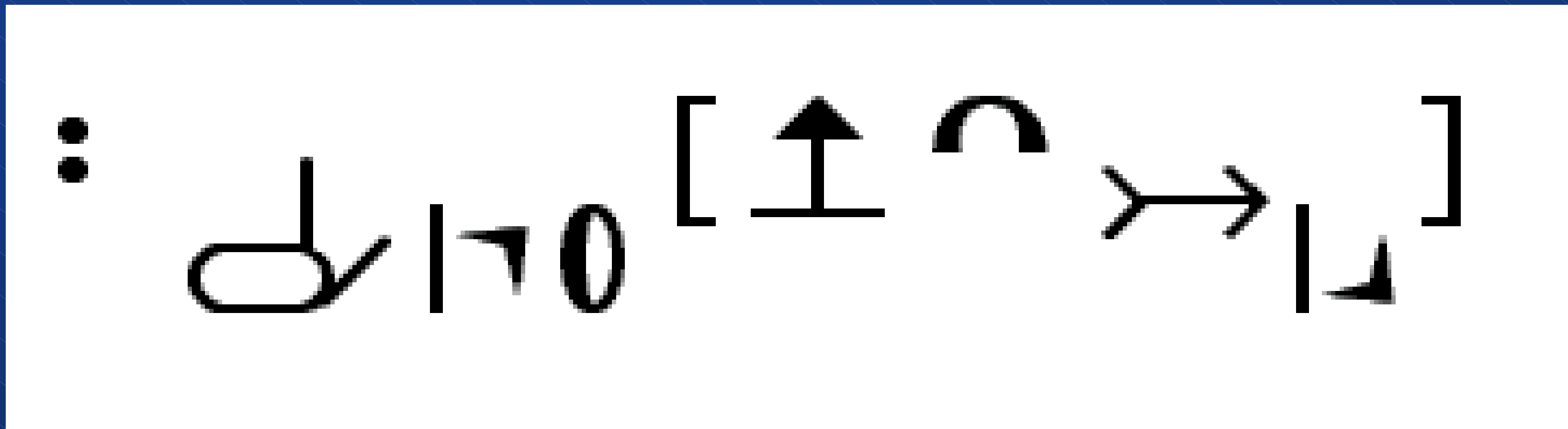
- *Straight*
- *Curved*
  - small, medium or large curvature of arc
- *Wavy and Zig-zag*
- *Circular and Elliptical*
  - varying no. of rotations
- *... All with varying direction/orientation*

Motion – Modality

- Fast
- Slow
- Rest – " Stoppage at start"
- Tense
- Sudden Halt

HamNoSys Example

DGS (German) Sign: “ GOING-TO”



XML

- *Represent Structured and "Semi-Structured" Data*
- *Textual Form*
  - tailored to transmission over WANs/Internet
- *An XML Document*
  - must be well-formed
  - may also be valid
    - *structure respects Document Type Definition – DTD (document may be "self-describing")*

XML Format

- *Use "nested labelled bracket" structure to delimit elements*
  - represent "brackets" by tags:

    <myelement ...> ... </myelement>
- *Element:*
  - may contain sub-elements and/or text
  - may have named attributes
- *DTD defines for each element type:*
  - content model
  - permitted attributes

Current SiGML Definition

- *Covers "Manual" subset of HamNoSys*
- *Embodied in SiGML DTD*
- *Two versions…*
- *"Initial" SiGML*
  - DTD as close as possible to HamNoSys
    - *rich in grammatical ambiguities…*
      - i.e. multiple ways of expressing the same thing
- *SiGML*
  - eliminates many of these ambiguities

DGS: “ GOING-TO”

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE sigml SYSTEM "sigmlv0.dtd">
<sigml>
    <avatar url="Simon.ava" id="A" alt="Simon"/>
    <sign gloss="GOING-TO">
    <!-- Taken from Hamnosys 2.0 manual, p.42, top line. -->
        <hamnosys_sign lr_symm="parallel">
            …
        </hamnosys_sign>
    </sign>
    …
</sigml>
```

“ GOING-TO” -contd.

```
<hamnosys_sign lr_symm="parallel">
    <handposture
        handshapeclass="ham_finger2"
        thumbpos = "ham_thumb_outmod"
        extfidir="direction_uo"
        palmor="direction_l">
    </handposture>
    <par_movement>
        <straightmovement
            direction="direction_o"
            curve="direction_u"
        />
        <handposture extfidir="direction_do"/>
    </par_movement>
</hamnosys_sign>
```

SiGML – Current State

- *Supporting tools*
  - translate from HamNoSys
  - use XSLT (for the second stage)
- *Definition – to come:*
  - non-manual enhancements
    - *more than HamNoSys*
  - multiple “ tiers”
    - *allow units bigger than a single sign*

Stages – Animation

1. NL (English)
2. Semantic Representation
3. Morphology (Sign-Language Specific)
4. Signing Gesture Notation (SiGML)
5. Animation

Animation

- *Pure Synthesis from SiGML is possible*
  - motion is “ robotic”
  - improve by use of appropriate non-linear interpolation
- *But Motion Capture gives authenticity*
  - Conjecture: Best result will come from a combination of purely synthetic and motion-captured elements.