Univ Access Inf Soc (2008) 6:405–418
DOI 10.1007/s10209-007-0094-8

LONG PAPER

A knowledge-based sign synthesis architecture

Stavroula-Evita Fotinea · Eleni Efthimiou · George Caridakis · Kostas Karpouzis

Published online: 20 November 2007
© Springer-Verlag 2007

Abstract This paper presents the modules that comprise a knowledge-based sign synthesis architecture for Greek sign language (GSL). Such systems combine natural language (NL) knowledge, machine translation (MT) techniques and avatar technology in order to allow for dynamic generation of sign utterances. The NL knowledge of the system consists of a sign lexicon and a set of GSL structure rules, and is exploited in the context of typical natural language processing (NLP) procedures, which involve syntactic parsing of linguistic input as well as structure and lexicon mapping according to standard MT practices. The coding on linguistic strings which are relevant to GSL provide instructions for the motion of a virtual signer that performs the corresponding signing sequences. Dynamic synthesis of GSL linguistic units is achieved by mapping written Greek structures to GSL, based on a computational grammar of GSL and a lexicon that contains lemmas coded as features of GSL phonology. This approach allows for robust conversion of written Greek to GSL, which is an essential prerequisite for access to e-content by the community of native GSL signers. The developed system is sublanguage oriented and performs satisfactorily as regards its linguistic coverage, allowing for easy extensibility to other language domains. However, its overall performance is subject to current well known MT limitations.

Keywords Sign synthesis · Linguistic knowledge · Lexicon and grammar coding · Virtual signing · Sign language resources

S.-E. Fotinea (✉) · E. Efthimiou
Institute for Language and Speech Processing,
Artemidos 6 & Epidavrou, Athens, Greece
e-mail: evita@ilsp.gr

E. Efthimiou
e-mail: eleni_e@ilsp.gr

G. Caridakis · K. Karpouzis
Image, Video and Multimedia Systems Lab,
9, Iroon Polytechniou St, Athens, Greece
e-mail: gcari@image.ntua.gr

K. Karpouzis
e-mail: kkarpou@cs.ntua.gr

1 Introduction

Since the issue of Universal Access to Information Society Technologies has first been raised, conventional computer-meditated human activities, as well as emerging services and applications, are influenced by the requirement for development of Information Society Technology products and services that are accessible to all citizens [34]. Adaptation-based techniques are already explored to cover the needs of universal accessibility of software user interfaces in respect to blind user populations [35]. Accessibility by deaf individuals implies the use of sign language, given that written text—opposite to a widely accepted misconception—does not provide an accessible means for the deaf. Worldwide surveys have proven that the reading capability of average deaf adults does not exceed the middle Primary School stage.

Consequently, designing any accessible system or tool for the Deaf requires integration of mechanisms that allow access to content by conveying meaning via 3D representation. Moreover, video—though it preserves naturalness of the signing utterance—is a rather static, not easily reusable source of linguistic content. Currently investigated

dynamic sign synthesis may well be considered as a viable alternative.

The approach to Greek sign language (GSL) synthesis presented in this paper is heavily based on experience gained from NLP applications of syntactic parsing and speech synthesis technologies for spoken languages. In GSL, as in any sign language, there is a closed set of phonological components [10, 15, 36, 38], various combinations of which generate every possible sign. Speech technology has exploited properties of phonological composition of words in orally uttered languages to develop speech synthesis tools for unrestricted text input. In the case of sign languages, a similar approach is experimented with in order to generate signs (i.e., word level linguistic units of sign languages) not by mere video recording, but rather by composition of sign phonology components. To achieve this, a library of sign notation features, among other linguistic primes, has been converted to motion parameters of a virtual agent (avatar). In order to extend the generative capacity of the system to phrase level, a set of core grammar rules provides structure patterns for GSL grammatical sentences, which may receive unrestricted word level signs on the leaves of the tree representations of the analysis frame.

To this end, the GSL NL knowledge of the conversion system consists of a lexicon annotated according to the Hamburg Notation System (HamNoSys) [20, 30] and a set of structure rules utilising strings of morphemes to compose core-signing utterances of GSL. HamNoSys provides a set of symbols for the phonological representation of signs.

Linguistic data to be signed are written Greek utterances. In order to handle written Greek input for conversion, a local statistical parser for Greek is used that outputs syntactic chunks on the basis of tag annotations on input word strings. The created chunks are next mapped to GSL structures, which provide the sign string patterns to be performed by the avatar. Mapping incorporates standard MT procedures to handle addition or deletion of non matching linguistic elements between the two languages, as well as to perform feature insertion on GSL heads, in order to provide for multilayer formation that characterises natural (complex) sign performance.

The paper is organised as follows. Section 2 discusses the coded GSL linguistic resources that are exploited in order to convert written Greek structures to GSL structures and to further visualise the signing utterance using a virtual signer (avatar). The sign lexicon, its multilayer enrichment, as well as the GSL structure rules, are also described. Section 3 presents the Greek to GSL converter, addressing the system's architecture, NLP procedures, implementation issues and coverage, while Sect. 4 discusses the virtual signing techniques that have been adopted for 3D sign generation, including implementation issues, manual feature performance and incorporation of nonmanual features. Section 5 presents the pilot system evaluation, while concluding remarks follow in Sect. 6.

2 GSL linguistic knowledge

Greek sign language synthesis [13] is heavily based on natural language (NL) knowledge, as is the case with other well tested sign synthesis systems [12]. This is necessary to guarantee, to an acceptable extent, the linguistic adequacy of the sign generation tool, given that data of signing corpora in the form of digital video were not annotated in a way that would allow an approach based on alternative methodologies, such as standard statistical processing of linguistic input.

In this respect, linguistic knowledge is exploited both in the GSL synthesis and the Greek to GSL conversion procedure.

This type of linguistic knowledge allows for robust conversion from written Greek text to GSL signing, resulting, in principle, in an application environment independent tool, which may support access by deaf users to any type of language e-content, if properly coded sublanguage tokens are available.

Furthermore, the GSL grammar makes use of the morpheme level as the principal structural unit for the construction of the system's grammar, based on a feature exploitation theoretical framework [3, 25, 27, 33].

Coding of GSL knowledge involves a lexicon of signs annotated as to the phonological composition of the lemmas, among other semantic and syntactic features, and a set of rules that allows structuring of core grammatical phenomena in GSL.

2.1 The sign lexicon

The system's lexicon contains sign lemmas described as to their phonological structure [9], i.e., the handshape for sign formation, hand movement, palm orientation and location in the signing space or on the signer's body. For the representation of the phonological features of GSL the extended HamNoSys notation system has been adopted. Every lemma appears in a list of default written Greek forms, where it is accompanied by the set of symbols which compose its HamNoSys string. The phonological structure of lemmas reveals a number of interesting parameters of sign formation as regards morpheme combinations for the creation of lexical items. For example, root morphemes of different semantic categories, such as the base signs for

'boy' and 'girl', provide for the formation of the signs 'brother' and 'sister' (token id 45 and 46 in Fig. 1) respectively, when combined with a semantic unit roughly interpreted as 'born by X'.

Decomposing the sign phonology allows for the development of an unrestricted, avatar-based device for sign generation [4, 21, 24, 37], which may compose a new sign, previously unknown to the system, as soon as it meets a string of symbols that dictate to an avatar a predefined sequence of motions.

The interesting point in respect to the adopted analysis is that the list of phonologically analysed items contains annotated strings which correspond to either simple or complex signs or to morphemes involved in sign formation [28].

Sign coding, except for symbols for motion, also involves coding of nonmanual features which obligatorily accompany hand action during articulation. Figure 2 demonstrates obligatory combination of manual features with the nonmanual mouthing feature in the articulation of the predicates 'run', 'scold', 'accuse' and 'kiss' (token id 358, 227, 185 and 379, respectively).

2.2 Multilayer enrichment

Multilayer phonological composition of signs makes use of a set of features, which, along with mouthing patterns, also incorporates features for facial expressions and body movement. This set includes eyebrows movement and eye tracking which both are significant parts of GSL sign formation (Fig. 16 below exhibits formation of the sign for 'children' involving eye gaze).

In Fig. 3, a sample of coding of the non manual features accompanying sign lemmas is presented. The 'yes' value dictates obligatory simultaneous performance with HamNoSys annotated hand motions (where 'no' dictates lack of

42	ΑΓΟΡΙ	
43	ΆΓΡΙΟΣ	
44	ΆΔΕΙΟΣ	
45	ΑΔΕΛΦΗ	
46	ΑΔΕΛΦΟΣ	
47	ΑΔΥΝΑΤΟΣ	
48	ΑΕΡΑΣ	
49	ΑΕΡΟΠΛΑΝΟ	
50	ΑΝΑΒΩ-ΦΩΣ	
51	ΆΝΕΜΟΣ	
52	ΑΝΗΦΟΡΑ	
53	ΑΝΘΡΩΠΙΝΗ-ΟΝΤΟΤΗΤΑ 2	
54	ΑΝΘΡΩΠΙΝΗ-ΟΝΤΟΤΗΤΑ-ΚΙΝΕΙΤΑΙ	
55	ΑΝΘΡΩΠΟΣ	
56	ΑΝΟΙΞΗ	
57	ΑΝΟΙΧΤΟΣ	

Fig. 1 HamNoSys annotated lemmas

ID	Lemma	HamNoSys	Mouthing
358	ΤΡΕΧΩ		OL7
227	ΜΑΛΩΝΩ		CN1
185	ΚΑΤΗΓΟΡΩ		CN17
379	ΦΙΛΩ		CN14

Fig. 2 HamNoSys annotated lemmas accompanied by information regarding obligatory nonmanual features

obligatoriness of feature and empty feature position declares that the specific feature is irrelevant). For instance, plural 'YOU' (token id 107) obligatorily requires eye tracking.

2.3 GSL structure rules

The set of rules of the GSL grammar module can handle sign phrase generation as regards the basic verb categories and their complements, as well as extended nominal formations.

The rules generate surface structures with a linear ordering that corresponds to basic sign sequences in a phrase. However, the maximal phrase level representations also contain features that provide linguistic information, which is expressed nonlinearly. The default case involves nonmanual information, arranged in a multilayer mode on structural heads or at the sentence level.

Typical instantiations of this are sentential negation and the presence of qualitative adjectives in nominal phrases. Negation is indicated by a complex nonmanual feature at sentence level, that has to be realised throughout performance of the predicative sign. Tokens like 'nice' or 'good' are formed by incorporating the adjectival value on the nominal morpheme by means of an appropriate mouth gesture. That is, for the oral string 'nice apple', for example, the GSL equivalent involves signing the head 'apple' while simultaneously performing the mouthing gesture that corresponds to the qualitative adjective ('nice'). To provide for proper representation and generation of such utterances, the GSL computational grammar makes use of condition dependent feature insertion rules, which apply on the output of morpheme based structure rules, where the later generate core signing utterance structures. The leaves of the so created structures are lemmas of the GSL lexicon. Lemmas and rules comprise the GSL coded knowledge that functions in a bidirectional way. It provides the linguistic descriptions that have to be represented by avatar motion, and also defines the output of the Greek to GSL conversion procedure.

Fig. 3 Nonmanual feature coding on GSL lexicon

ID	lemma	word family	lation for or	default nm	word class	an be a topi	gaze follo	verb modif	umber plur	epeat plura	hand plura	rative aspe	inutive' as	tense aspe	peat aspe	actic mov
1	ΑΡΙΘΜΟΣ	numbers	No	No	No			No	No		Yes		No	No	Yes	Ye
2	ΕΒΔΟΜΑΔΑ		No	No	Yes			No	Yes		No		No	No	Yes	N
3	ΚΟΚΚΙΝΟ		No	No	No			No	No		No		No	No	No	N
4	ΑΓΑΠΗ	αισθήματα	No	No	No			Yes	No		No		No	Yes	No	N
5	ΑΓΑΠΩ	αισθήματα	No	No	No			Yes	No		No		Yes	Yes	No	N
6	Αγελάδα	βοοειδή	No	No	Yes			No	Yes		No		No	No	No	N
7	ΑΓΟΡΙ	αρσενικά	No	No	Yes			No	Yes		No		Yes	Yes	No	N
8	Άγριος	σκύλος	No	Yes	Yes			Yes	Yes		No	Yes	Yes	Yes	Yes	N
9	Άδειος		No	No	No			No	No		No		No	No	No	N
10	ΑΔΕΛΦΗ		No	No	No			No	No		No		No	No	No	N
11	ΑΔΕΛΦΟΣ		No	No	No			No	No		No		No	No	No	N
12	Αδύνατος		No	Yes	No			No	No		No		No	No	No	N
13	Αέρας	αέρας	No	No	No			Yes	No		No		No	No	No	N
14	ΑΕΡΟΠΛΑΝΟ	ταξινομητής	No	No	No	Yes		No	No		No		No	No	No	N
15	ΑΝΑΒΩ-ΦΩΣ	φως	Yes	Yes	No			Yes	No		No		No	No	No	N
16	Άνεμος	αέρας	No	No	No			Yes	No		No		No	No	No	N
17	ΑΝΗΦΟΡΑ	ταξινομητής	Yes	No	No			Yes	No		No		No	No	No	N
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--
107	ΕΣΕΙΣ	δείξη	No	No	No	yes		No	No		No		No	No	No	N
108	ΕΣΥ	δείξη	No	No	No			No	No		No		No	No	No	N
109	ΕΧΘΡΟΣ		No	No	No			Yes	No		No		No	No	No	N
110	ΖΕΣΤΗ 1	καλοκαίρι	No	Yes	No			Yes	No		No		No	No	No	N
111	ΖΕΣΤΗ 2	δροσιά	No	No	No			Yes	No		No		No	No	No	N
112	ΖΕΣΤΟΣ (ΚΑΥΤΟΣ)	καίει	No	Yes	No			Yes	No		No		No	No	No	N
113	ΖΥΓΑΡΙΑ	ζυγαριά	No	No	No			Yes	No		No		No	No	No	N
114	ΖΥΓΙΖΩ	ζυγαριά	No	No	No			Yes	No		No		No	No	No	N
115	Ζώο	βοοειδή	No	No	No			No	No		No		No	No	No	N
116	Ήλιος	φως	No	No	No			Yes	No		No		No	No	No	N
117	ΗΜΕΡΑ		No	No	No			No	No		No		No	No	No	N
119	ΗΜΕΡΟΛΟΓΙΟ		No	No	No			No	No		No		No	No	No	N
120	Ήμερος	ήσυχος	No	No	No			Yes	No		No		No	No	No	N
121	ΘΑ / ΜΕΛΛΟΝ	εκτός	Yes	No	No	yes		No	No		No		No	No	No	N
122	Θάλασσα	ταξινομητής επιφάνειας	No	No	No			Yes	No		No		No	No	No	N

3 Greek to GSL converter

3.1 System architecture

The Greek to GSL conversion tool consists of two submodules, the Greek to GSL mapping and the GSL sign synthesis, as schematically depicted in Fig. 4 [14]. Input to the conversion procedure are written Greek sentences parsed to chunks. The conversion tool incorporates transfer-based MT mechanisms, where the source language is written Greek and the target language is GSL.

The system may currently handle successfully a subset of GSL grammar as regards both linear and nonlinear grammar aspects (see Sect. 3.3.4, below). In the context of a specific sublanguage, there is full vocabulary coverage, while the overall conversion performance is satisfactory. The system is subject to well known MT limitations, performance results raising significantly within a well defined sublanguage. Portability of the system to new language pairs is foreseen. The system being transfer based though, restructuring of the mapping submodule as well as a proper transfer dictionary are required, in order to capture mapping demands put by the new language pair. The Greek to GSL converter, along with the related database for coded linguistic knowledge, are currently working off line. Incorporation of the system to an Internet application under development, which supports avatar performance within an educational platform [31, 32], foresees that the MT component runs as a package on users' workstations.

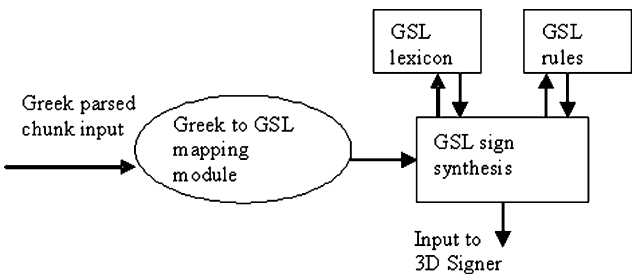

Fig. 4 Schematic presentation of the Greek to GSL conversion tool

3.2 NLP procedures

3.2.1 Greek to GSL mapping module

Written Greek sentences (text) are processed by a shallow, statistical parser/chunker [2], which also makes use of linguistic information based on morphological tags on words of phrasal strings. Parsing results in structured chunks, which correspond to grammatically adequate syntactic units of the Greek language with feature values for morphological annotations of input words and structural annotation for phrases. The parser for Greek has been developed at a previous stage, independently of the Greek to GSL conversion tool that is subject to the present discussion, and it was intended to handle free language input in written form. Exploitation of the specific parser in the conversion tool was decided on the ground that it was an already available, well tested system for shallow parsing, which may handle successfully large amounts of natural language data. This fact allows for unrestricted text (e-content) handling with respect to written language input. Incorporation of the parser into the Greek to GSL converter provides source language analysis in a representation that can be input to the mapping submodule, but it can not alone

increase adequate conversion capacity, which is currently restricted to domain-specific linguistic utterances, as also mentioned in Sect. 3.1 above.

The Greek to GSL Mapping module transfers the written Greek chunks to equivalent GSL structures, and aligns input tagged words with corresponding signs or features on sign heads.

Various mapping operations perform addition, deletion or swap of utterance constituents, as well as feature insertion. An example of sentence level mapping handles predicates with empty pronominal subject of Greek, generating a double deictic pronoun subject in GSL. Under this rule, sentential strings as '*τρώω*' (I-EAT) are mapped to a GSL structure that results to the string 'I-EAT-I' which is the grammatical option of the language for the construction of the predicate "eat". The rule that maps chunks is presented in Fig. 5, where the chunks of the verb group (vg) on the left-hand side are the output of the shallow parser for Greek and the corresponding verb group for GSL is indicated on the right-hand side. The rule generates the positions for the deictic pronoun that serves as subject and has to be signed by the virtual signer in order to result in a grammatically acceptable signed utterance in GSL.

Another example involves the Greek noun phrase '*ωραίο μήλο*' (nice apple), that has to match the GSL structure where the specific noun phrase has to be realised as a complex sign by performing the manual sign for the nominal head ('apple') simultaneously with the mouthing gesture for 'nice'. In the mapping module, the adjective chunk is replaced by the corresponding mouthing feature. This procedure is combined with a general mapping rule that makes use of the semantic tag 'qualitative' on adjective heads and deletes the input chunk related to the adjective word, while it creates a corresponding feature on the nominal head. This feature may receive several values deriving from mapping between specific adjectives and mouthing gestures.

SYN [cl	SYN [cl
SYN [vg	SYN [Pr_deictic
SYN *sing	LU *εγώ* *Pr01Sg*
TOK *τρώω* W	SYN /Pr_nm]
Pr01Sg	SYN [vg
SYN *sing	SYN *sing
SYN /vg]	LU *τρώω* W
SYN [*XP]	*Pr01Sg* vb_sg
SYN /cl]	SYN [Pr_deictic
	LU *εγώ* *Pr01Sg*
	SYN /Pr_nm]
	SYN *sing
	SYN /vg]
	SYN [*XP]
	SYN /cl]

Fig. 5 Empty pronominal subject mapping with doubled deictic pronoun

3.2.2 *Rule-based GSL synthesis*

Rule-based GSL synthesis is responsible for converting the GSL mapping output of the previous stage (e.g., right hand side of the mapping rule of Fig. 5 above) to sequences of commands to be performed by the 3D Sign Generation module, namely the VRML model performing the signs.

The rule-based GSL synthesis module, containing all structural representations that describe the rules, which generate core GSL sentences, interacts with both the GSL lexicon and the library of features which define avatar motion under different conditions. In the case of the example 'nice apple' described above, the chunk description provides information related to signing the nominal token, while in parallel adding the mouthing gesture that corresponds to each quantitative modifier (adjective). In order for the virtual signer to perform this example, the module reads the corresponding HamNoSys notation as well as the mouthing gesture from the library.

3.3 Implementation issues

3.3.1 *General issues*

The rule sets responsible for the mapping between Greek and GSL structures are programmed in Java to allow for quick and efficient application development compatible with all system platforms (Windows 95/NT/XP, Linux and Solaris 2.3 and later). Extensible markup language (XML) technology has been utilised as a means for describing structured documents in a reusable format (deriving from SGML—ISO 8879). XML advantages, on the one hand, are found in management and flexibility during communication/information exchange in a multilayer processing of annotated corpora. Java technology, on the other hand, contains embedded tools (object/class) for the management of XML texts that can be utilised to allow for simple, quick and efficient text handling. Hence, the conversion tool utilises multilevel XML-based annotated sentences.

3.3.2 *Methodological approach*

The conversion tool performs top–down, rule-based metasyntactic analysis, its input being parsed Greek text in a multilevel XML annotated corpus form. The input text is processed via a set of rules that perform matching between chunked Greek sentences and GSL structures. The rules are

organised in three sets, the structure set, the chunk set and the feature set. The structure rule set allows for linguistic actions involving (conditioned) reordering of chunk sequences (swap, or 3 and 4-elements reordering) to reflect the morpheme order of GSL (similar to word order in spoken languages). A second set of rules performs on the chunk level, allowing for (conditioned) addition, deletion or modification of specific entries. A third set of rules applies to feature bundles, by inserting GSL-specific features, while deleting or modifying features irrelevant to GSL synthesis.

In the current configuration, the grammar writer (user) can arrange the rules into rule sets and consequently execute one specific rule, all rule sets or any batch defined by the user. Rule execution is iterative and for each iteration all specified rules are examined, the output of each rule serving as the input to the next one, provided that the rule context is satisfied.

3.3.3 GSL synthesis environment

The structure of an XML rule contains an 'if-part' to control the linguistic context in which the rule applies and a 'then-part' to describe the linguistic actions to be taken (an example is given in Fig. 6). The label 'flag' controls the presence (or absence) of a feature in the input chunk, the labels 'tag', and 'taglist' the matching criteria with GSL grammar requirements, while the labels 'lemma' and 'lemmalist' check the item or the group of items (e.g., with the same semantic or morphosyntactic information) relevant to the rule.

The 'then-part' of the rule defines labels such as 'chunk type', the type of 'action' to be performed, as well as new values ('newLemma', 'newTag', 'newType', 'GslFeature') to appear in the output.

The currently available types of actions include:

- «addition» when a new lemma is added
- «modification» when the current lemma is modified with respect to some of the values of its feature bundle
- "deletion" where the current lemma is totally omitted
- "copy" of the current lemma along with all its features
- "2 position reordering" if two chunks are being swapped
- "3 position reordering" applied in the form of "A B C becomes A C B" (or any other required ordering).

Figure 6, presents a sample of the coded rule that adds a double deictic pronoun before and after the verb "eat" or any other verb coded in the list "food". In this list, predicates such as "drink", "eat", "swallow", etc., are coded; these predicates present the same function in GSL structures, namely appearing surrounded by deictic pronouns.

Finally, in Fig. 7, a screen shot of the application environment is depicted, where the upper half part of the

Fig. 6 Rule example (id 6) for the case of the double deictic pronouns and related GSL feature

```
<!--IF-PART --!>
<IF rule="6" type="*">
<!—A verb that is member of the group "food" (search is separate file) is sought. --!>
        <ATTR flag="+"  lemma="*" lemmaList="false" tag="*" taglist="false"></ATTR>
</IF>
<IF rule="6" type="vg">
        <ATTR flag="+"  lemma="food" lemmaList="true" tag="*" tagList="false"></ATTR>
</IF>
<IF rule="6" type="*">
 <ATTR flag="+" lemma="*" lemmaList="false" tag="*" taglist="false"></ATTR>
</IF>

<!--THEN-PART --!>
<THEN rule="6" type="*" action="*" newLemma="" newTag="" newType = "" GslFeature="">
<ATTR flag="+" lemma="*" lemmaList="false" tag="*" taglist="false"></ATTR>
</THEN>
<THEN rule="6" type="vg" action="+" newLemma="εγώ" newTag="Pr01Sg" newType =
"Pr_deictic" GslFeature="CN14">
<!--ACTION='+', namely adds deictic Pr_deictic and special GSL feature before Verb "vg" --!>
        <ATTR flag="+" lemma="food" lemmaList="true" tag="*" taglist="false"></ATTR>
</THEN>
<THEN rule="6" type="*" action="+" newLemma="εγώ" newTag="Pr01Sg" newType =
"Pr_deictic" GslFeature="CN14">
<!--ACTION='+', namely adds deictic Pr_deictic and special GSL feature before "*", namely
after Verb "vg" --!>

<ATTR flag="+" lemma="*" lemmaList="false" tag="*" taglist="false"></ATTR>
</THEN>
<THEN rule="6" type="*" action="*" newLemma="" newTag="" newType = "" GslFeature="">
<!—ACTION ='*', namely copies everything until the end of the sentence !-->
        <ATTR flag="+" lemma="*" lemmaList="false" tag="*" taglist="false"></ATTR>
</THEN>
```

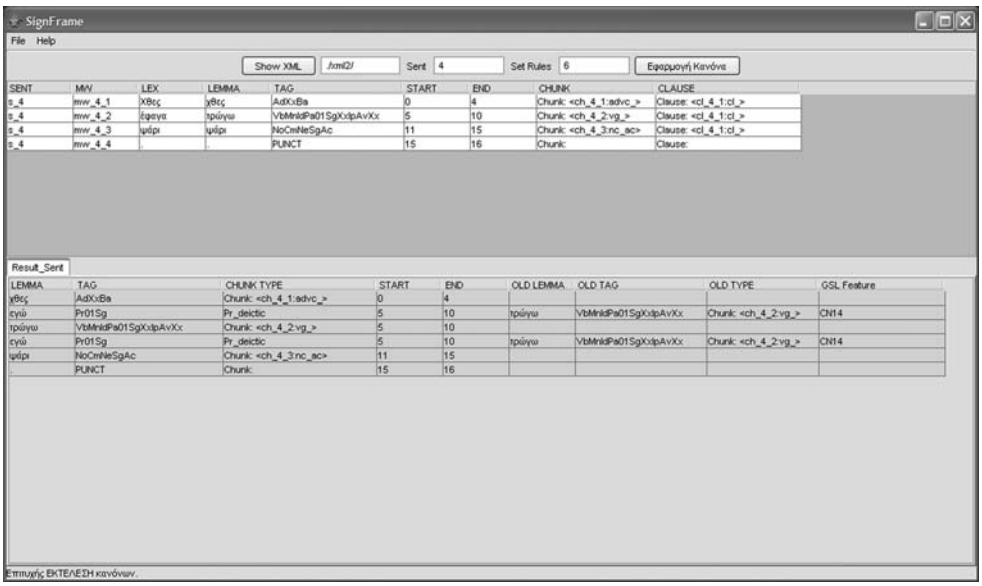

Fig. 7 The application environment of the conversion utility and the execution of the rule id 6 (of Fig. 6) on the input sentence "Χθες έφαγα ψάρι (I ate fish yesterday)". The output is depicted on the bottom half of the screenshot

screen contains a description of the sentence to be converted and the rule/rule set to be used, while the bottom half of the screen is reserved for demonstration of rule execution results and application internal messages.

3.3.4 Current grammar coverage

The computational grammar of GSL currently handles various phrase-level phenomena which involve both linear and multilayer articulation mechanisms [7]. Grammar rules provide structures enriched with GSL-specific features. Such features support, for example, plural formation, where noun phrase (NP) plural values result from agreement checking inside the NP. Similarly, the semantic values related to aspect declaration in GSL are handled as features denoting language intrinsic adverbial properties such as continuation, duration, degrading, intensity or repetition.

The structure rules adopted to construct the conversion output are based on theoretical linguistic analysis of language data [6].

As regards predicate classification, empirical evidence and related analysis support three main clusters: 'Simple Predicates', 'Predicates of Direction' and 'Spatial Predicates'. The current grammar implements a pattern which incorporates both simple and spatial predicate formations. Predicates of direction are not yet treated, since they heavily involve the use of classifiers that are not yet implemented in the conversion grammar.

Implementation has adopted the word order Agent–Complement–Predicate, supported by theoretical analysis as the basic word order of the language. It also allows for an adequate handling of the set of phenomena that take place on clause level and involve sentential negation, tense declaration and interrogation and emphasis assigned to either predicate arguments or various sentential adjuncts (i.e., temporal adverbs).

Implementation of NP conversion mainly involves constituent arrangement around the nominal head, including actions of deletion of information irrelevant to articulation in 3D space (i.e., determiner deletion), and feature insertion obligatory for the reconstruction of information articulated in a multilayer manner in GSL (i.e., mouthing patterns parallel to movement in sign formation for quantitative adjectives).

As regards conversion coverage, however, many GSL specific phenomena still remain unsolved. The next research target involves integration of classifiers in structure formation, efficient handling of the signing space, and discourse modelling.

4 3D sign generation

4.1 Virtual character animation

In order to produce reusable animation sequences, the development team investigated a number of virtual character (VC) technologies. A major requirement was the ability to animate the VC via a scriptable, human-understandable language that supports at least minimal language structures, such as grouping of animation commands, to be executed in parallel or in sequence, and caters for the timing of the execution of each command. Language structures would facilitate script reuse and dynamic production of derivative word forms (e.g., plural forms), while making scripting authoring a straightforward process for

animation designers by eliminating the need for a specialised front-end; timing is also of great importance, since changing the speed of a specific movement can introduce important nonverbal characteristics, such as size or magnitude of the described concept (e.g., run faster). Besides scripting, an important requirement was also the absence of proprietary software to be installed in the client computer; that is expected to make adoption of the platform a straightforward task and minimise setup requirements. The Web3D technologies adopted [16] consist of a VRML, h-anim compatible VC, controlled by scripts written in the STEP language (Scripting Technology for Embodied Persona) [17], animated via a Java applet and illustrated via a standard plug-in in a web browser. Figure 8 illustrates the required transformations for the right hand to assume the "d"-handshape. The same code of the left hand can be compiled by mirroring the described motion, while other, more complicated handshapes can start with this representation and subsequently introduce the extra components into it.

The h-anim standard [19] has proposed a virtual character infrastructure, called h-anim figure, in which the human body is modelled by a number of segments (such as the forearm, hand and foot), connected to each other by joints (such as the elbow, wrist and ankle—see Fig. 9). In this framework, a human body is defined as a hierarchy of segments and articulated at joints; relative dimensions are proposed by the standard, but are not enforced, permitting the definition and animation of cartoon-like characters. Another attribute is that prominent feature points on the human body are defined in a consistent manner, via their names and actual locations in the skeleton definition. As a result, a script or application that animates an h-anim compatible virtual character (VC) is able to locate these points easily and concentrate on the high level appearance of the animation process, without having to worry about the actual 3D points or axes for the individual transformations. In the developed architecture, this is of utmost importance, because sign description is performed with respect to these prominent positions on and around the virtual signer's body. Moreover, the h-anim standard provides a systematic approach to representing humanoid models in a 3D graphics and multimedia environment, where each humanoid is abstractly modelled in terms of structure as an

```
par([
turn(humanoid,r_thumb1,rotation(1.9,1,1.4,0.6)
,very_fast),
turn(humanoid,r_thumb2,rotation(1,0.4,2.2,0.8)
,very_fast),
turn(humanoid,r_thumb3,rotation(1.4,0,0.2,0.4)
,very_fast),
...
])
```

Fig. 8 STEP code for a handshape

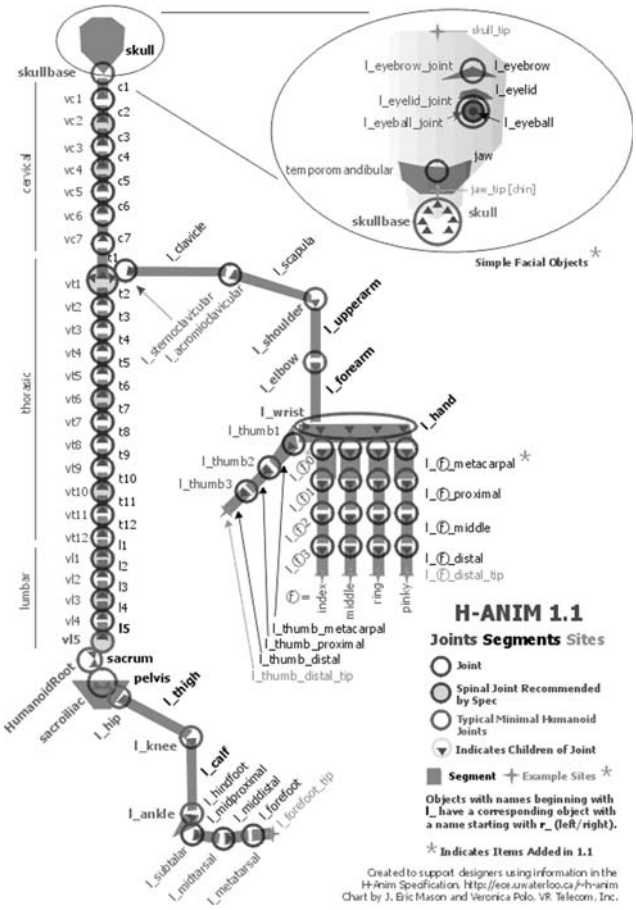

Fig. 9 h-anim skeleton infrastructure

articulated character, embedded and animated using the facilities provided by the selected representation system. Hence, the h-anim standard defines animation as a functional behaviour of time-based, interactive 3D, multimedia formally structured characters, leaving the particular geometry definition in the hands of the modeller/animator.

In the 3D Generation module, the STEP language provides the interaction level between the end user and the signing subsystem. The major advantage of this choice was the dissociation of the scripting language and the definition of the geometry and hierarchy of the VC. This dissociation results in the reusability and scalability of the scripting code without the need to remodel the VC. As shown next, the fact that the script is human-understandable caters for easy reuse and expansion, for example, when a single sign can be part of another sign (e.g., the sign for 'bull' is used as part for 'cow'). A similar case involves mirroring of handshapes, as for example in the case of the sign for 'donkey', or the formation of plural form by repetition of the same movement pattern in neighbouring point in the signing space (e.g., the sign for 'children').

While the STEP language possesses many useful characteristics, it essentially is a research effort, resulting in

minimal adoption in commercial applications, also lacking support of facial expressions which are essential to convey multilayer information.

Therefore, the keyframe-based animation designed in the STEP platform was transformed in a series of atomic rotations per actual animation frame, encoded in text files compatible with the MPEG-4 SNHC standard [29]. This transformation is performed by interpolating the evolution of rotations over the number of requested frames, which is resolved by the timing parameter of the script; for example, if the script states that a body part rotation must be performed in 0.5 s, this translates into 12 frames of a 25-frame-per-second animation.

Adopting the MPEG-4 standard allows for dissemination of the system towards both the academia and the industry. The created animation files still possess the interoperability of script-based animation, and can therefore be reused in a number of compatible software products, even on devices with low computational capabilities and hence, benefit from other concepts defined within the MPEG-4 standard, such as streaming over a telecommunications network. In order to play back the designed animation files, the Greta MPEG-4 player [5] is used, which is able to support higher-level animation editing and features such as emotions or composite gestures.

4.2 Manual features performance

In the case of GSL sign performance, HamNoSys annotated input has to be decoded and transformed to sequences of scripted commands. A demo site of the current performance of the system can be found online at http://www.image.ece.ntua.gr/~gcari/gslv (the VC shown is "yt", by Matthew T. Beitler, available at http://www.cis.upenn.edu/~beitler). Figure 10 shows the VC signing the GSL sign for "child", while Fig. 11 shows an instance of the plural formation of the same sign. The design of the automated script production system, combined with the related plural formation rule for GSL accompanying HamNoSys lemma annotation, enables using the default sign in order to construct its plural form. In the case of Fig. 10, plural formation involves repetition of the basic sign with simultaneous hand sliding to the signer's right. The sliding direction, along with the required secondary movement, is incorporated in the HamNoSys annotation for the relevant lemma.

Plural formation in GSL makes use of a set of rules, the application of which is appropriately marked in the lexical database in respect to each lemma. A different plural formation instantiation is provided in Fig. 13.

In Fig. 12, the VC performs the GSL sign for "day", while in Fig. 13 its numerical plural form "two days" is exhibited. In this case, different coding in lexicon results in

Fig. 10 The GSL sign for "child"

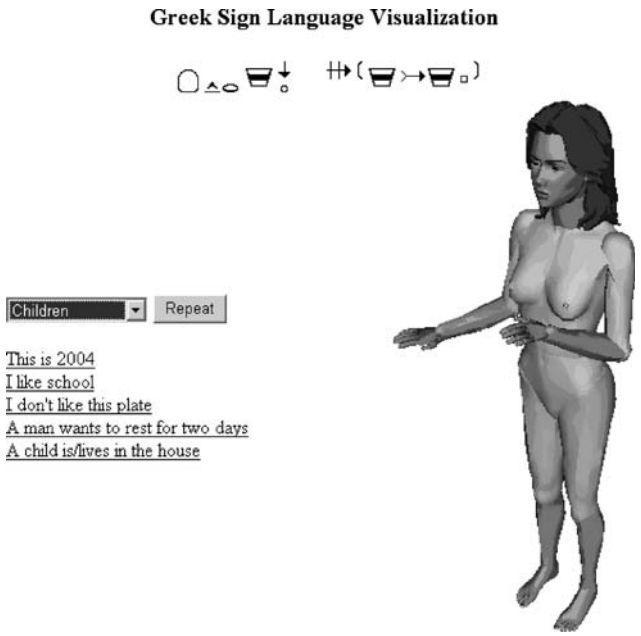

Fig. 11 The GSL sign for "children"

the appropriate VC performance, where a two-finger handshape is used to perform the basic sign movement, instead of the default straight-index finger handshape.

In Fig. 13, the VC is used in a frontal view to demonstrate the corresponding property of Blaxxun Contact 5 (VRML plug in) [1], which allows for better perception of this specific sign detail. Despite the default tilted view being the one of choice from the part of the users, the ability to show frontal and side views of a sign is crucial, since it caters for displaying the differences between

Fig. 12 The GSL sign for "day"

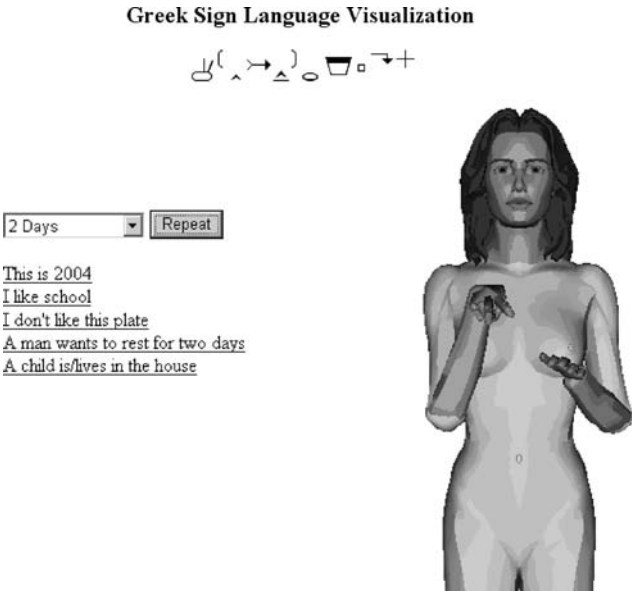

Fig. 13 The frontal view of "two days"

similar signs and brings out the spatial characteristics of signs [23, 24].

4.3 Nonmanual features incorporation

4.3.1 Head movement and eye gaze

When discussing NL knowledge of the system, special reference was made to nonmanual features of GSL, which are obligatory elements of sign formation [11], in many cases functioning as the differentiating features between otherwise identical sign formations. These features compose the multilayer information, which has to be processed in parallel with linearly ordered basic manual sign components, in order to achieve sign performance as close to natural as possible.

Among nonmanual features, head movement and eye gaze are of significant importance, as they convey specific grammatical meaning on word or phrase level (i.e., negation, verb declination, sentential tense, role in discourse, etc.) and typically follow the hand movement trace. Implementation of these features significantly increases the degree of acceptance of the performed sign by natural signers.

Head movement is widely used in discourse situations where by default the signer faces his/her interlocutor and has to use (different) positions in the signing space to place the human entities involved in narration. Hence, when a third person is included in the plot of the narration, the signer's head and gaze are turned towards his/her specified position, so as to indicate reference to events related to this person. In order to change subject of reference, the same pattern is applied turning towards the position of the newly involved individual.

In Fig. 14, a narration example is shown, where two individuals—other than the interlocutor(s)—are involved. In the picture pair (a) and (b), the signer conveys information related to the first person (X) not being present, where in (a) the signer positions X in the signing space and in (b) he conveys the content of X's action indicated by the turn of the head towards X's position. A similar situation is presented in picture pair (c) and (d), where change of direction of the head signifies reference to the second individual (Z) involved in the same narration.

Grammatical information realised via head movement and eye gaze, and coded in the GSL grammar module, allows for synthesis of utterances by the avatar with minimal technical cost. For example, since temporal relations are expressed by different eye gaze positions, the avatar may assign sentential tense to the utterances it composes by exploiting the relevant features whenever they are present in the output of the Greek to GSL conversion procedure.

The issue of eye gaze following the hand movement track during sign animation was tackled as a combination of rotating vectors about an arbitrary axis and standard forward kinematics [26]. Thus, given the rotation axis and the relevant angle at the shoulder and elbow joints, one can readily calculate the 3D position of the wrist joint. Then, this position is calculated in relation to the position of the "skullbase" to provide the "look_at" vector for the virtual signer's head, using the following steps [18](Figure 15), since the signer is looking straight ahead:

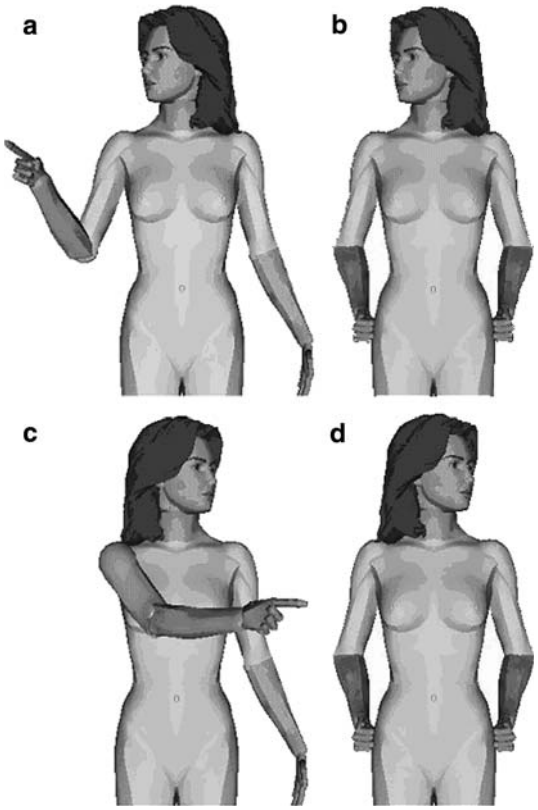

Fig. 14 **a** Signer positioning individual X in signing space, **b** signer signing "X sits down", **c** signer positioning individual Z in signing space, **d** signer signing "Z sits down"

$$N_{current} = [0 \quad 0 \quad 1] \quad N_{target} = \frac{P_{wrist} - P_{skullbase}}{|P_{wrist} - P_{skullbase}|}$$

$$Axis = N_{target} \times N_{current}$$

$$Angle = -\arccos(N_{target} \cdot N_{current})$$

Eye gaze is one of the obligatory nonmanual features participating in word level sign formation, and its implementation significantly improves naturalness of avatar performance. In order to incorporate eye gaze in the VC's performance, the system recognises the relevant feature accompanying basic manual descriptions in the sign lexicon database. Figure 16 shows performance of the sign for "children" with incorporated eye gaze feature effect.

4.3.2 *Facial expressions*

Facial expressions, usually referred to as nonmanual grammar markers, nonmanual behaviors, or nonmanual signals, are an important part of sign languages, since they can alter the meaning of a sign. Facial expressions are rule-governed, indicating, e.g., YES/NO questions, which are differentiated from facial expressions used to formulate questions: who, why, when, where, etc. Facial expressions

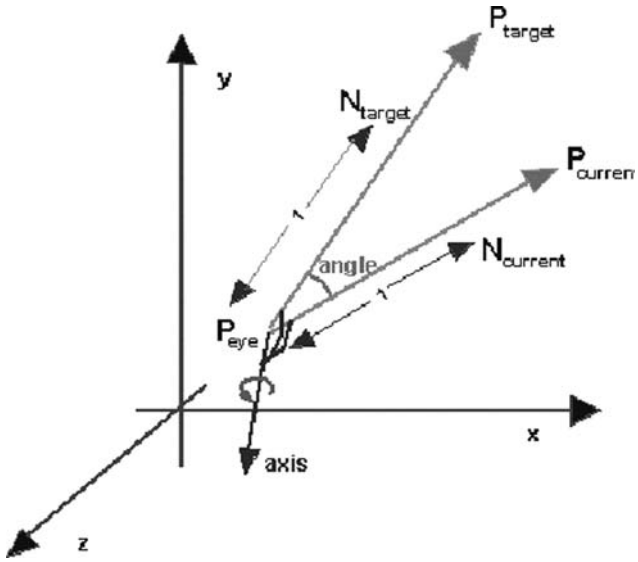

Fig. 15 Overview of vectors and angles used in eye gazing (EuclideanSpace URL)

Fig. 16 Eye gaze performance when signing "children"

are also used in combination with signs and fingerspelled messages to communicate specific vocabulary, intensity, and subtleties of meaning [22]. These actions add meaning to what is being signed, much like vocal tones and inflection add meaning to spoken utterances by means of prosody, where indicative functions of prosody are:

- to delimit syntactic and semantic units within an utterance
- to indicate focus in an utterance
- to convey pragmatic notions such as illocutionary force
- to convey nuances of meaning.

In this context, lack of facial expressions in a sign generation engine would correspond to the absence of an

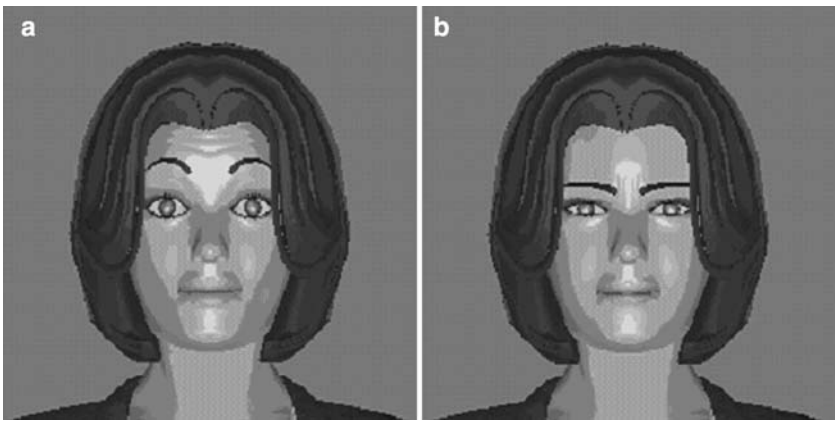

Fig. 17 Eyebrow effect for wh-question formation (**a**) and indication of emphasis (**b**)

NLP system's capacity to handle questions. As far as conveyance of feeling, interest or focus is concerned, the human signer may exhibit a range of emotion from sadness to excitement, depending on the subject matter of the signing.

Given that the HamNoSys notation set used for GSL lemma descriptions does not extend to facial expressions, manual intervention was necessary in order to reduce the effect of the lack of rule-based implementation at the level of sign synthesis. On the basis of information expressed by specific grammar features on input strings, injection of sign synthesis information, relevant to a specific facial expression, takes place in between the components of a phrase or in parallel to the manual execution of a single sign, in order to augment grammatical correctness and semantic completeness of the GSL synthesis output. Figure 17 depicts the effect of implementing eyebrow movement for wh-question formation (a) and indication of emphasis (b).

5 Pilot system evaluation

The first test bed for GSL sign synthesis was provided in the domain of e-education. A signing virtual human tutor (avatar) was incorporated in an Internet based prototype educational platform, targeted to the population of Greek deaf pupils of primary schools, in order to support the GSL grammar lesson [31, 32]. Evaluation was performed on both technical and functional aspects of the platform. Different user groups participated in the different stages of evaluation. Continuous internal evaluation has been carried out by experts in the areas of sign language and linguistics, IT technology, language teaching and deaf education. End user evaluation was carried out by native GSL signers, both students and their tutors. It took place in three circles and was organised by the Hellenic Federation of the Deaf, which also hosted evaluation sessions. A pilot evaluation procedure was carried out on an early prototype version of the platform in an experimental environment. Two further evaluation procedures used revised versions of the platform, modified according to the feedback gained in previous evaluation sessions. Subject to evaluation were the usability of the system, as well as its appeal to the users with respect to navigation, educational targets and acceptability of the VC signer. Users' guided responses and free comments were coded in categories.

Evaluators' comments in respect to the virtual signer involved naturalness, accuracy of performance when signing specific signs, its appearance and the point of viewing the avatar on the screen, as well as suggestions for zooming in the hands and change of body (signing background) colors to increase comprehensibility of the performed signs. More specifically, the avatar was in general found to perform well enough (=comprehensively enough [16]) but to be somehow "cold blooded". To improve friendliness, one suggestion was to give it the characteristics of a girl rather than of an adult woman. A general preference was noticed for the rest position to lie the hands beside the body.

As regards performance of specific signs, there were comments related to articulation accuracy, e.g., finger stretching when performing a specific handshape, width of motion, co-ordination of hands in two hand signs, or ambiguous sign articulation. Lack of naturalness is a generally identified problem in comments on the abrupt avatar motion, in comparison to smoothness of transitions in human signing.

6 Conclusion

The combination of linguistic knowledge and avatar performance described in this paper allows for dynamic conversion from written Greek text to GSL away from restrictions put by the use of video. Furthermore, the adopted analysis of GSL allows for handling multilayer information, which is part of the obligatory set of features

which have to be realised for a grammatical GSL utterance to be performed. The resulting tool exploits animation technologies, along with electronic linguistic resources, and constitutes a sign generation mechanism adaptable to various environments [8], also addressing the demand for universal access to e-content.

A number of technical issues still remain open with respect to animation technologies, and it may be true that avatar representations will hardly ever reach the quality of representation of natural signing by video display. However, virtual signing seems to be the only solution for unrestricted sign generation against the problem of e-content accessibility, and it can also perform successfully enough in specific sublanguage applications, an across the board well-known situation in relation to NLP performance.

The ultimate challenge, though, remains handling of unlimited linguistic data in MT conditions. It is still too difficult to produce acceptable sentences in the context of automatic translation of unrestricted input for any language pair. This procedure becomes even more difficult in the case of a less researched language with no written tradition such as GSL. Realistically, the teams involved in the reported research may expect as an optimum result the successful use of automatic translation in a restricted, sublanguage oriented environment with predetermined semantic and syntactic characteristics.

Acknowledgments The authors wish to acknowledge the assistance of all groups involved in the implementation of the work presented here, especially the GSL research group of ILSP and the IVMS lab. This work was partially funded by the national project grants SYNNENOESE (GSRT: eLearning 44).

References

1. Blaxxun Contact 5. http://www.blaxxun.com/home/index.php?option=com_content&task=view&id=42&Itemid=85
2. Boutsis S., Prokopidis P., Giouli V., Piperidis S.: A robust parser for unrestricted Greek text. In: Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC 2000, Athens, pp. 467–473 (2000)
3. Carpenter B.: The Logic of Typed Feature Structures. Cambridge University Press, London (1992)
4. Chen, Y., Gao, W., Fang, G., Wang, Z., Yang, C., Jiang, D.: Text to Avatar in Multi-modal Human Computer Interface. In: Proceedings of Asia-Pacific CHI (APCHI2002) (2002)
5. De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B.: From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Int. J. Hum. Comput. Stud. 59 (1–2), 81–118 (2003)
6. Efthimiou, E.: Processing cumulative morphology information in GSL: the case of pronominal reference in a three-dimensional morphological system. In honor of G. Babiniotis, National and Kapodestrian University of Athens (2006, in press)
7. Efthimiou, E., Fotinea, S.-E., Sapountzaki, G.: Processing linguistic data for GSL structure representation. In: Proceedings of the Workshop on the Representation and Processing of Sign Languages: Lexicographic matters and didactic scenarios, Satellite Workshop to LREC-2006, Genoa, pp.49–54 (2006)
8. Efthimiou E., Sapountzaki G., Karpouzis K., Fotinea S.-E.: Developing an e-learning platform for the Greek sign language. In: Miesenberger K., Klaus J., Zagler W., Burger D. (eds) Lecture Notes in Computer Science, vol. 3118, 1107–1113. Springer, Heidelberg (2004)
9. Efthimiou, E., Vacalopoulou, A., Fotinea, S.-E., Steinhauer, G.: Multipurpose design and creation of GSL dictionaries. In: Proceedings of the Workshop on the Representation and Processing of Sign Languages "From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication", Satellite Workshop to LREC-2004 Conference, Lisbon, pp. 51–58 (2004)
10. Efthimiou, E., Katsoyannou, M.: Research issues on GSL: a study of vocabulary and lexicon creation (in Greek). In: Studies in Greek Linguistics, Computational Linguistics, vol. 2, pp. 42–50, Aristotle University of Thessaloniki, Thessaloniki (2001)
11. 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) Designing a More Inclusive World, pp. 127–136, Springer (2004)
12. Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I.: Development of language processing support for the ViSiCAST project. In: Proceedings ASSETS 2000, 4th International ACM SIGCAPH Conference on Assistive Technologies, Washington DC (2000)
13. Fotinea, S.-E., Efthimiou, E., Karpouzis, K., Caridakis, G.: Dynamic GSL synthesis to support access to e-content. In: Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction (UAHCI 2005), Las Vegas, Nevada, USA (2005)
14. Fotinea, S.-E., Efthimiou, E., Kouremenos, D.: Generating linguistic content for Greek to GSL conversion. In: Proceedings of the HERCMA-2005 Conference (The 7th Hellenic European Conference on Computer Mathematics and its Applications), Athens (2005)
15. Fischer, S.D., Gough, B.: Verbs in American sign language. Sign Lang Stud **18**, 17–48 (1978)
16. Foulds, R.: Biomechanical and perceptual constraints on the bandwidth requirements of sign language. IEEE Trans Neural Syst Rehab Eng **12**, 65–72 (2004)
17. Huang, Z., Eliens, A., Visser, C.: STEP: A scripting language for embodied agents. In: Proceedings of the Workshop on Lifelike Animated Agents (2002)
18. EuclideanSpace http://www.euclideanspace.com/maths/algebra/vectors/lookat/index.htm
19. Information technology—Computer graphics and image processing—Humanoid animation (H-Anim) International Standard (ISO/IEC FCD 19774—Humanoid animation (H-Anim)):http://www.h-anim.org/Specifications/H-Anim200x/ISO_IEC_FCD_19774/
20. Hanke, T.: HamNoSys in a sign language generation context. In: Schulmeister, R., Reinitzer, H. (eds) Progress in sign language research. In honor of Siegmund Prillwitz / Fortschritte in der Gebärdensprachforschung. Festschrift für Siegmund Prillwitz. (International Studies on Sign Language and Communication of the Deaf; 40) Hamburg, 249–264 (2002)
21. Karpouzis, K., Caridakis, G., Fotinea, S.-E., Efthimiou, E.: Educational resources and implementation of a Greek sign language synthesis architecture. In: Computers & Education, Web3D technologies in learning, education and training, vol. 49(1), pp. 54–74, Elsevier, Amsterdam (2007)
22. Karpouzis, K., Raouzaiou, A., Drosopoulos, A., Ioannou, S., Balomenos, T., Tsapatsoulis, N., Kollias, S.: Facial expression and gesture analysis for emotionally-rich man-machine

interaction. In: Sarris, N., Strintzis, M. (eds) 3D Modeling and Animation: Synthesis and Analysis Techniques, Idea Group, Toronto, pp. 175–200 (2004)
23. Kennaway, R.: Experience with, and requirements for, a gesture description language for synthetic animation. In: Proceedings of the 5th International Workshop on Gesture and Sign Language based Human-Computer Interaction, Genova (2003)
24. Kennaway, R.: Synthetic animation of deaf signing gestures. In: Proceedings of the International Gesture Workshop, London (2001)
25. Koskenniemi, K.: Finite-state parsing and disambiguation. In: Proceedings of the 13th International Conference on Computational Linguistics (ACL), Helsinki, pp. 229–232 (1990)
26. Lengyel, E.: Mathematics for 3D Game Programming and Computer Graphics, 2nd edn. Charles River Media, Hingham Massachusetts (2003)
27. Lenseigne, B., Dalle, P.: A computational model for sign language grammar. In: 2nd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Sciences and Linguistics, Poland (2005)
28. Padden, C.: Interaction of Morphology and Syntax in American Sign Language. Garland, New York (1988)
29. Preda, M., Prêteux, F.: Critic review on MPEG-4 face and body animation. In: Proceedings IEEE International Conference on Image Processing (ICIP 2002), Rochester, pp. 505–508 (2002)
30. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide. Broschur (ISBN 3-927731-01-3) Signum Verlag, Seedorf (1989)
31. Sapountzaki, G., Efthimiou, E., Fotinea, S.-E.: Digital technology in Greek sign language teaching in primary school curriculum (in Greek). In: Proceedings of Hellenic Conference on Digital Educational Material: Issues of creation, exploitation and evaluation, Volos, 6–7 April 2006 (in press)
32. Sapountzaki, G., Efthimiou, E., Fotinea, S.-E.: e-Education in the Greek Sign Language using an Internet based platform (in Greek). In: Proceedings of Workshop: e-Learning using ICT, Lamia, 26–28 May 2006 (in press)
33. Shieber, S.M.: Constraint-Based Grammar Formalisms. MIT Press, Cambridge (1992)
34. Stephanidis, C.: Editorial. Int J Universal Access Inf Soc **1**, 1–3 (2001)
35. Stephanidis, C., Savidis, A.: Universal access in the information society: methods, tools, and interaction technologies. Int J Universal Access Inf Soc **1**, 40–55 (2001)
36. Stokoe, W., Kuschel, R.: A Field Guide for Sign Language Research. Linstock Press, Maryland (1979)
37. Suszczanska, N., Szmal, P., Francik, J.: Translating Polish texts into sign language in the TGT system. In: Proceedings of 20th IASTED International Multi-Conference Applied Informatics, Innsbruck, pp 282–287 (2002)
38. Sutton-Spence R., Woll B.: The linguistics of BSL—an introduction. Cambridge University Press, London (1999)