

Dynamic GSL synthesis to support access to e-content

Stavroula - Evita Fotinea¹, Eleni Efthimiou¹, Kostas Karpouzis², George Caridakis²

¹Institute for Language and Speech Processing, Artemidos 6 & Epidavrou, Athens, Greece

²Image, Video and Multimedia Systems Lab 9, Iroon Polytechniou Str., Athens, Greece

Email: evita@ilsp.gr, eleni_e@ilsp.gr, kkar pou@cs.ntua.gr, gcari@image.ntua.gr

Abstract

This paper presents the modules that comprise a sign generation tool for Greek Sign Language (GSL). The tool combines Natural Language (NL) knowledge with avatar technology in order to allow for unrestricted generation of sign utterances. The NL knowledge of the system is exploited in the context of typical NLP procedures, which involve syntactic parsing of linguistic strings as well as structure and lexicon mapping according to standard Machine Translation (MT) practices. The coding on linguistic strings which are relevant to GSL, provide instructions for the motion of a Virtual Reality Modeling Language (VRML) signer, that performs the corresponding sign streams. Dynamic synthesis of GSL linguistic units, is achieved by combining syntactic rules of the language with a lexicon that contains lemmas coded as to features of GSL phonology. This approach, although sublanguage oriented as regards our current implementation, allows for unrestricted conversion capacity of written Greek to GSL, which is an essential prerequisite for access to e-content by the community of native GSL signers.

1 Introduction

The here presented approach to GSL synthesis 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 close set of phonological components (Fischer & Gough, 1979, Stokoe & Kuschel, 1978, Sutton-Spence & Woll, 1999, Efthimiou & Katsoyannou, 2001), various combinations of which generate every possible sign. Speech technology has exploited properties of phonological composition of words in respect to aural 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 (=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 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 sign units on the leave level.

To this end, the GSL NL knowledge of the conversion system consists of a lexicon annotated according to the Hamburg Notation System (HamNoSys) and a set of structure rules utilizing strings of morphemes to compose core signing utterances of GSL.

Linguistic data to be interpreted to signs are written Greek utterances. In order to handle written Greek input for conversion, we utilise a local statistical parser for Greek that outputs syntactic chunks on the basis of tag annotations on input word strings. The so 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.

2 NL Knowledge

GSL synthesis is heavily based on Natural Language (NL) knowledge of the system. This is necessary to guarantee, to an acceptable extent, the linguistic adequacy of the sign generation tool. In this respect, linguistic knowledge is exploited both in the GSL generation 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, to a tool independent of application environment, that may support access by deaf users to e-content.

Furthermore, GSL grammar makes use of the morpheme category as the principal structural unit, instead of lexical categories used in traditional rule based electronic grammar sets (Shieber, 1992, Koskenniemi, 1990) as is the case with most known spoken languages.

2.1 GSL knowledge

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.

2.1.1 The sign lexicon

The system's lexicon contains sign lemmas described as to their phonological structure (Efthimiou et al., 2004b), 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 annotation system (Prillwitz et al., 1989) 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 (Figure 1 provides a demonstration of HamNoSys annotated strings, among which various sign formation properties are to be identified, as, for example, the combination of root morphemes of different semantic categories for the formation of the signs for 'brother' and 'sister' (id 45, 46) respectively, where a semantic unit roughly interpreted as 'born by X' is combined with the signs for 'boy' and 'girl' respectively). Decomposing the sign phonology allows for the development of an unrestricted, avatar based device for sign generation (Karpouzis et al., 2004), 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 our adopted analysis, is that the list of phonologically analyzed items, contains annotated strings which correspond to either simple or complex signs or to morphemes involved in sign formation. Sign coding is further enriched to provide, except for HamNoSys symbols for motion, other non-manual obligatory features, which accompany hand action in order to make linguistic content fully acceptable by the native GSL signers' community.

In Figure 2, a part of the lexicon is shown, where HamNoSys annotated lemmas are accompanied by information on obligatory non-manual features (if necessary). In the case of this example, coding involves various mouthing gestures, as for instance, for the verbs 'τρέχω' (run), 'μαλώνω' (scold), 'κατηγορώ' (accuse), 'φιλώ' (kiss).

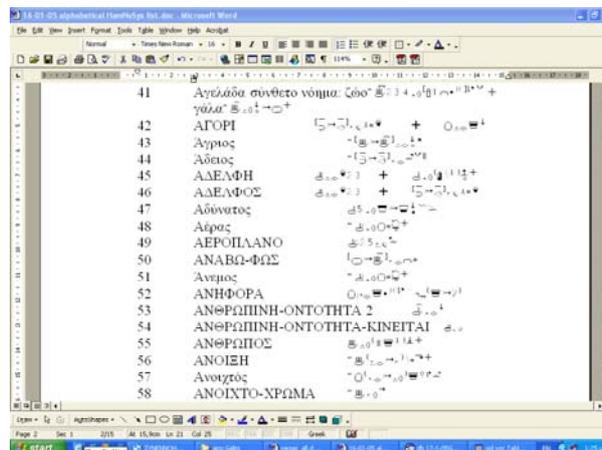


Figure 1: HamNoSys annotated lemmas

is expressed non-linearly. The default case involves non-manual 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 non-manual feature on sentence level, that has to be realized throughout performance of the verbal head sign. Adjectives like ‘nice/good’ are formed by incorporating the adjectival value on the nominal head morpheme by means of an appropriate mouth gesture. For example, for the aural string ‘nice apple’, the GSL equivalent involves signing the head ‘apple’ while simultaneously performing the mouthing gesture that corresponds to the qualitative adjective (‘nice’).

As already mentioned, GSL structure rules utilize strings of morphemes to compose core signing utterances, enriched with multilayer information. 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.

2.2 Greek to GSL converter

The Greek to GSL conversion tool consists of three submodules, the Shallow Parsing for Greek, the Greek to GSL Mapping and the GSL Synthesis. The Greek to GSL conversion tool is schematically depicted in Figure 4.

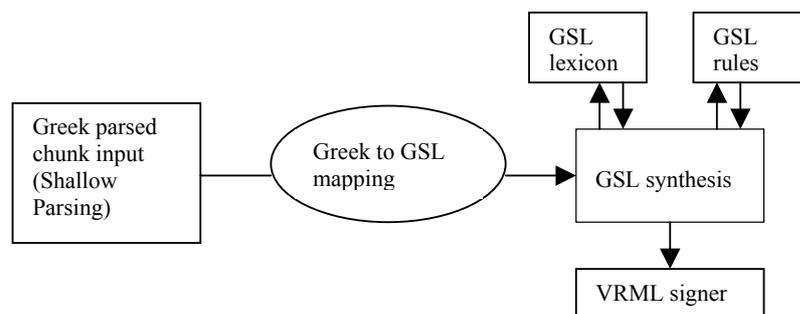


Figure 4: Schematic presentation of the Greek to GSL conversion tool

In the Shallow Parsing submodule, Greek written sentences (text) are processed by a shallow, statistical parser (Boutsis et al., 2000) 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 Greek language with feature values for morphosyntactic annotations on input words and structural annotation on phrases. Incorporation of the described procedure to the conversion tool, takes advantage of shallow parsing potential to successfully handle large amounts of natural language data. This fact allows for, in principle, unrestricted handling of text in the form of e-content. The chunks created by the parser serve as the input of the Greek to GSL mapping module. When GSL reaches a stage of analysis that can successfully cover the major language performance capacity, the conversion tool may allow for unrestricted generative capacity given the limitations that generally hold in respect to NLP performance. However, our present target is acceptable performance in well defined sublanguage environments.

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. For instance, the Greek noun phrase ‘ωπαίο μήλο’ (nice apple) has to match the GSL structure needs where the specific noun phrase has to be realized 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.

interactive 3D, multimedia formally structured characters, leaving the particular geometry definition in the hands of the modeler/ 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 re-usability and scalability of the scripting code without the need to remodel the VC in any sense.

3.1 Manual features performance

In our case, the HamNoSys annotated GSL input has to be decoded and transformed to sequences of scripted commands. A demo site of the current performance of our 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>). To demonstrate productivity of the adopted engine, we will use the example of GSL plural formation. Plural formation in GSL makes use of a set of rules, the application of which is appropriately marked in our lexical database in respect to each lemma. Figure 6 shows the VC signing the GSL sign for “child”, while Figure 7 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 lemma HamNoSys annotation, enables use of the default sign in order to construct its plural form. In the case of of the example of “child”, 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 are incorporated in HamNoSys annotation for the relevant lemma. A different plural formation instantiation is provided in Figure 9. In Figure 8 the VC performs the GSL sign for “day”, while in Figure 9 its numerical plural form “two days” is exhibited. In this case, different coding in lexicon results in 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.



Figure 6: The GSL sign for “child”



Figure 7: The GSL sign for “children”



Figure 8: The GSL sign for “day”



Figure 9: The frontal view of “two days”

In Figure 9 the VC is used in a frontal view to demonstrate the corresponding property of Blaxxun Contact 5 (VRML plug in), 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 similar signs and brings out the spatial characteristics of signs (Kennaway, 2001, 2003).

3.2 Non-manual features incorporation

When discussing NL knowledge of the system, special reference was made to non-manual features of GSL which are obligatory elements of sign formation, very often also functioning as the differentiating features between otherwise identical sign formations. These features compose the multilayer information which has to be processed in parallel with basic phonological sign components, in order to grammatically perform sign formation.

Among non-manual features, head movement and eye gaze are of significant importance. When encountered, they convey specific grammatical meanings on word or phrase level (i.e. negation, verb declination, sentential tense, role in discourse, ...) and they 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 signing space to place the person or persons 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 the specified position, so as to indicate reference of events related to this person. In order to make reference to another person, the same pattern is applied turning towards the position of the new person involved.

In Figure 10, a narration example is shown, where two persons –other than the interlocutor(s)- are involved. In the picture pair (a) and (b), the signer conveys information related to the first person (John) not being present, where in (a) the signer positions John in the signing space and in (b) he conveys the content of John's action indicated by the turn of the head towards John'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 person (Mary) involved in the same narration.

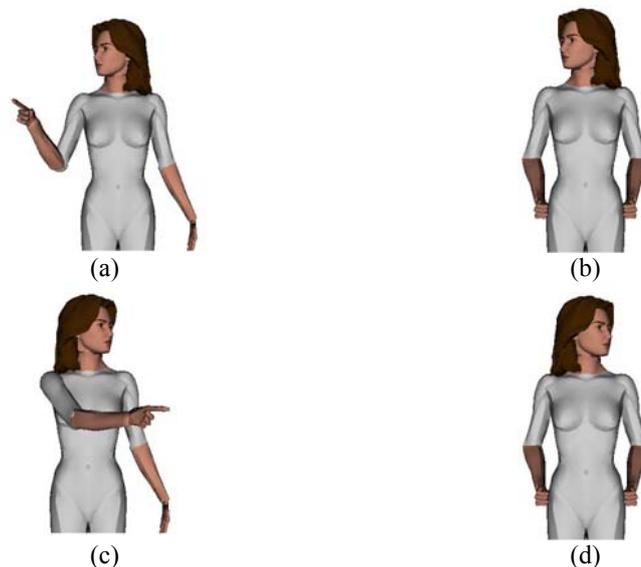


Figure 10: (a) Signer positioning John in signing space, (b) Signer signing “John sits down”, (c) Signer positioning Mary in signing space, (d) Signer signing “Mary sits down”

Grammatical information realised via head movement and eye gaze, and being 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 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 (Lengyel, 2003). 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 (EuclideanSpace URL, Figure 11):

since the signer is looking straight ahead,

$$N_{\text{current}} = [0 \ 0 \ 1]$$

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

$$N_{\text{target}} = \frac{P_{\text{wrist}} - P_{\text{skullbase}}}{|P_{\text{wrist}} - P_{\text{skullbase}}|}$$

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

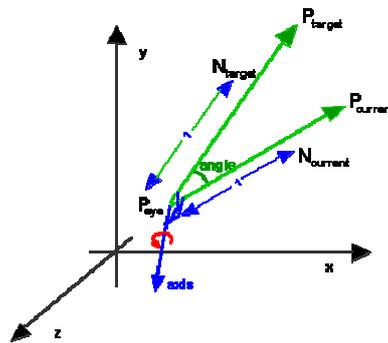


Figure 11: Overview of vectors and angles used in eye gazing (EuclideanSpace URL)

Eye gaze is one of the obligatory non-manual features participating in word level sign formation, implementation of which 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 phonologic descriptions in the sign database. Figure 12 shows performance of the sign for “children” with incorporated eye gaze feature effect.



Figure 12: Eye gaze performance when signing “children”

3.3 Problems and limitations

The most important technical problems include a solution for smooth transition between concurrent signs and fusion between handshapes so that neighboring signs in a sentence appear as naturally articulated as possible. This issue has been tackled using an interesting feature of the STEP engine, which at any time can return the setup of the kinematic chain for each arm. As a result, when the sign that is next in a sequence begins, the kinematic chain is transformed to the required position without having to take into account its setup in the final position of the previous sign. In general, this would be problematic in general purpose animation, since the h-anim standard itself does not impose any kinematics constraints; thus, random motion might result in physiologically impossible, puppet-like

animation. In the case of signing though, almost all action takes place in the signing space in front of the signer and starting from the head down to the abdomen; in this context, there are no abrupt changes in the chain setup.

4 Conclusion

The combination of linguistic knowledge and avatar performance described above, allows for dynamic conversion from written Greek text to GSL away from restrictions put by the use of video. Furthermore, our adopted analysis of GSL allows for handling the multilayer information which is part of the obligatory set of features which have to be realized 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 (Efthimiou et al., 2004a) also answering the demand for Universal Access to e-content.

A number of technical issues still remain open in 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 here reported research, may expect as an optimum result the successful use of automatic translation in a restricted, sub-language oriented environment with predetermined semantic and syntactic characteristics.

Acknowledgements

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 SYNNEOESE (GSRT: eLearning 44).

References

- Blaxxun Contact 5, <http://www.blaxxun.com/en/products/contact/>
- Boutsis, S., Prokopidis, P., Giouli, V. & Piperidis, S. (2000). A Robust Parser for Unrestricted Greek Text. In *Proc. of the 2nd International Conference on Language Resources and Evaluation, LREC 2000*, Athens, 467-473.
- Efthimiou, E., Sapountzaki, G., Karpouzis, K., Fotinea, S-E. (2004a). 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 (LNCS)*, Springer, Vol. 3118, 1107-1113.
- Efthimiou, E., Vacalopoulou, A., Fotinea, S-E., Steinhauer, G. (2004b). Multipurpose Design and Creation of GSL Dictionaries. In *Proc. 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, 30 May 2004, Lisbon, Portugal, 51-58.
- Efthimiou, E. & Katsoyannou, M. (2001). Research issues on GSL: a study of vocabulary and lexicon creation. In *Studies in Greek Linguistics*, Vol. 2, Computational Linguistics, 42-50 (in Greek).
- Fischer, S.D., Gough, B. (1979). Verbs in American Sign Language. In *The signs of language*. Klima, E.S. and Bellugi, U. (Eds.), Cambridge, Mass; London : Harvard Univ. Pr.
- Foulds, R. (2004). Biomechanical and Perceptual Constraints on the Bandwidth Requirements of Sign Language, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, March 2004, 65-72.

Huang, Z., Eliens, A., and Visser, C. (2002). STEP: A Scripting Language for Embodied Agents, In *Proc. of the Workshop on Lifelike Animated Agents*.

EuclideanSpace URL: <http://www.euclideanspace.com/maths/algebra/vectors/lookat/index.htm>

ISO/IEC 19774 — Humanoid Animation (H-Anim), <http://www.h-anim.org>

Karpouzis, K., Caridakis, G., Fotinea, S-E., Efthimiou, E. (2004). Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture. In *Proc. of the First International Workshop on Web3D Technologies in Learning, Education and Training (LET-WEB3D)*, September 30 – October 1, 2004, Udine, Italy, 8-15.

Kennaway, R. (2003). Experience with, and Requirements for, a Gesture Description Language for Synthetic Animation. In *Proc. of the 5th International Workshop on Gesture and Sign Language based Human-Computer Interaction*, Genova.

Kennaway, R. (2001). Synthetic Animation of Deaf Signing Gestures. In *Proc. of the International Gesture Workshop*, City University, London.

Koskenniemi, K. (1990). Finite-state parsing and disambiguation. In *Proc. of the 13th International Conference on Computational Linguistics (ACL)*, Helsinki, 229-232.

Lengyel, E. (2003). *Mathematics for 3D Game Programming and Computer Graphics*, Second Edition, Publisher: Charles River Media.

Prillwitz et al. (1989). HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide. Broschur / Paperback (ISBN 3-927731-01-3).

Shieber, S.M. (1992). *Constraint-Based Grammar Formalisms*. MIT Press, Cambridge, Massachusetts.

Stokoe, W. & Kuschel, R. (1978). *For Sign Language Research*, Linstock Press

Sutton-Spence, R. & Woll, B. (1999). *The Linguistics of BSL- an introduction*. Cambridge Univ. Press