A dynamic environment for Greek Sign Language Synthesis using virtual characters

G. Caridakis, K. Karpouzis Image, Video and Multimedia Systems Lab 9, Iroon Polytechniou Str.

> Athens, Greece +30 210 7724352

{gcari, kkarpou}@image.ece.ntua.gr

ABSTRACT

In this paper, we present a system that performs sign language synthesis in the framework of an educational platform for young deaf children. The proposed architecture is based on standardized virtual character animation concepts for the synthesis of sign sequences and lexicon-grammatical processing of Greek sign language (GSL) sequences. A major advantage of the proposed architecture is that it goes beyond the usual single-word approach which is linguistically incorrect, to provide tools to dynamically construct new sign representations from similar ones. Words and phrases are being processed and the resulting notation subset of a lexical database, HamNoSys (Hamburg Notation System), eventually transformed into GSL and animated on the clients' side via an h-anim compliant avatar.

Categories and Subject Descriptors

J.5 [Arts and Humanities]: Language translation, H.5.1 [information interfaces and presentation]: Multimedia Information Systems– Artificial, augmented, and virtual realities, H.5.2 [information interfaces and presentation]: Multimedia Information Systems– Training, help, and documentation

General Terms

Design, Human Factors, Standardization, Languages.

Keywords

Sign language synthesis, linguistic content, h-anim, educational resources.

1. INTRODUCTION

Greek Sign Language (GSL) is a natural visual language used by the members of the Greek Deaf Community with several thousands of native or non-native signers. Research on the grammar of GSL per se is limited; some work has been done on individual aspects of its syntax, as well as on applied and educational linguistics. Comparison of core vocabulary lists exhibit many similarities with sign languages of neighboring countries, while in morphosyntax GSL shares the same crosslinguistic tendencies as many other well analyzed sign languages [1][13].

GSL has developed in a social and linguistic context similar to most other sign languages. It is used widely in the Greek deaf community but also by a large number of hearing non-native signers of GSL, mainly students of GSL and families of deaf people. Although the exact number of hearing students of GSL in Greece is unknown, records of the Greek Federation of the Deaf G. Sapountzaki, S-E. Fotinea, E. Efthimiou

Institute for Language and Speech Processing Artemidos 6 & Epidavrou

> Athens, Greece +30 210 6875300

{eleni_e, evita}@ilsp.gr

(GFD) show that, in the year 2003 about 300 people were registered for classes of GSL as a second language. The recent increase of mainstreamed deaf students in education, as well as the population of deaf students scattered in other institutions, minor town units for the deaf and private tuition may well double the total number of secondary and potential sign language users. Official settings where GSL is being used include 11 Deaf clubs in Greek urban centers and a total of 14 Deaf primary, secondary and tertiary educational settings.

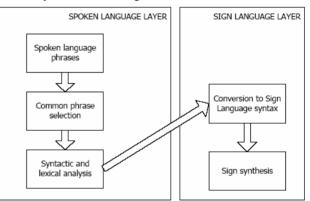


Figure 1. Overview of the proposed architecture

In consultancy with the Greek Pedagogical Institute, the SYNENNOESE project helps young pupils acquire the proper linguistic background so that they can take full advantage of the new accessible educational material. The platform offers students the possibility of systematic and structured learning of GSL for either self-tutoring or participation to virtual classroom sessions of asynchronous teaching, and its design is compatible with the principles that generally define systems of open and distant learning. Besides teaching GSL as a first language, in its present form the platform can be used for the learning of written Greek through GSL, and it will also be open to future applications in areas of other subjects in the school curriculum.

Figure 1 describes the abstract architecture and dataflow between the components of the integrated system. In this paper we describe the procedures followed during the compilation of the educational material and the implementation of the sign language synthesis component of the educational platform. In this process we utilized existing software components for the web-based animation of an h-anim virtual character [9]; the adoption of widely accepted character definition and animation standards caters for the extensibility and reusability of the system resources and its content.

2. LINGUISTIC AND EDUCATIONAL RESOURCES OF THE PROJECT

The linguistic part of the project is based on overall assumptions for the adequacy of signed languages as by Stokoe [16] and Woll and Kyle [11], among many. Greek sign language is analyzed to its linear and non-linear (simultaneous) components [14]. The linear part of the language involves any sequences of lexical and functional tokens and their syntactic relations, while non-linear structures in GSL, as in all known sign languages, are present in all levels of the grammar. Each sign in GSL is described as to its handshape, location, movement, orientation, number of hands and use of any obligatory non-manually articulated elements (e.g. mouth patterns, head and shoulder movements, facial expression and other non-manual features), based on the Stokoe model.

Many of the grammar rules of GSL are derived from the analysis of a digital corpus that has been created by videotaping native signers in a discussion situation or when performing a narration. This procedure is required, because there exists little previous analysis of GSL and rule extraction has to be based on actual data productions of native signers. The basic design of the system, except for the educational content this currently supports, focuses on the ability to generate sign phrases, which respect the GSL grammar rules in a degree of accuracy that allows them to be recognized by native signers as correct utterances of the language.

In this respect the SYNENNOESE project offers a great challenge for in-depth work on both directions, lexicography and linguistic analysis of GSL. For the first time, research goes beyond a mere collection of glosses and moves further from many previous bilingual dictionaries of sign languages [3], into the domain of productive lexicon [17], i.e. the possibility of building new GSL glosses following known structural rules, and also challenge automatic translation in predictable environments, using an effective module/interface for the matching of structural patterns between the written input and the signed output of the platform.

At the stage of grammatical analysis, findings from other sign language grammars, as well as the views of our deaf native user consultants are taken into account in order to verify findings. Tools utilized for the transcription and notation include HamNoSys, a pictographic notation system developed by the University of Hamburg for the description of the phonology of signs [15]. This notation forms the corpus of GSL lemmas while for the representation of sequential structures, i.e. in the phrase level, the ELAN language annotator developed by the Max-Planck Institute of Psycholinguistics in Nijmegen, the Netherlands, will be used. We considered these two systems as most suitable to the text-to-sign animation according to reviews of recent relevant projects.

The test bed learning procedure concerns teaching of GSL grammar to early primary school pupils, whereas the platform also incorporates a subsystem that allows approach by the deaf learner to material available only in written Greek form by means of a signed summary. The learning process in practice will involve an initiator of the session, the students in groups or alone and a teacher-facilitator of the process, physically present with the students. The process can take place in real-time or can be relayed. There is provision of a virtual white-board, icon banks and chat board visible in the screen along with the virtual signer for common use in the classroom. The participants will also be

able to see each other in real time through a web camera, in order to verify results of GSL learning.

The first group of exercises deals with signs that use the same handshape but start from different positions with respect to the signer's body or the neutral signing space and consist of different movements. An example of such a group in GSL includes the words "table", "house", "donkey", "slipper" and "tent". In this framework, young pupils are initially presented with the VC signing each word in a particular group and a sketch depicting the same concept; the use of sketches instead of written words is adopted since very young pupils have not developed skills related with spoken or written languages and thus, their mother tongue is the relevant sign language. In the following, pupils go through a number of drills, similar to the ones found in usual language teaching classes. These drills consist of choosing the correct sketch relating to a random sign performed by the VC and matching different instances of the VC with the correct sketch, by picking from an on-screen sketch pool.

The second group of exercises includes signs with similar or semantically related meaning, signed with the same or different handshapes. An example is the group "human", "tall", "fat", "child" and "female". The drills here are the same with the ones in the first exercise group, as is also the case with the third group of exercises. In this category, sign pairs are formed, consisting of signs composed of same phonological features (handshape, movement, location, palm orientation) but differing in their grammatical classification, e.g. "sit-chair", "eat-food" and "love verb-love noun" by means of movement repetition.

3. TECHNICAL CONSIDERATIONS

The developing team has adopted Web 3D technologies which have surpassed traditional approaches such as video and still images. A VRML, h-anim compatible model is controlled by the STEP engine. Such an approach provides many advantages. The user / student can examine the gestures from many different points of view. Actually the default viewpoint has been selected in such a way so as to make the gesture clear for inspection as to all the dimensions in 3D space. This variety in viewpoints is essential to the learning procedure of sign languages in general and GSL in particular. Accessibility is another feature very handy when using a platform for teaching sign language to public schools. The centralized control of such architecture adds reusability and changeability to the platform. The effectiveness of this approach can be easily demonstrated considering a remote, isolated school needing to be coherent with all the other schools using the platform. One can only imagine how fatiguing such a task would be if a standalone application model was adapted. Since a web approach was to be followed, one could argue that, video streaming would be an acceptable solution. The bandwidth and storage load is an obvious advantage of Web 3D solutions, which combined with the reusability of such an approach, makes it clearly advantageous. Additionally we reviewed some educational features of these approaches. It has been proven that the transduction process and the persona effect have beneficial impact in education [5]. During the learning procedure the student reacts positively to the presence of a human tutor, physical or virtual. Finally transduction, the process that converts information into forms conceivable to us (motion in the case of sign language), is

very important since the perception of animation is the primary means of retention.

3.1 Sign language synthesis

Different approaches to the problem of sign language visualization have been applied in several implementations. Fingerspelling of an animated hand is obviously the simplest solution. Of course adopting such a simplified approach expunges many features of the sign language. Actually rarely a signer finger spells when signing. A more sophisticated approach suggests of an avatar, not compliant with the h-anim standard, can animate whole words without including eye gazing, facial expressions or animation of whole phrases. Additionally many similar projects have the form of stand alone applications. This lack of accessibility, via web, downgrades the concept from an educational point of view. On the other hand some notable features are very challenging and would consider incorporating them in our platform. Such features are user dynamic gesture production, web browser plug-in concept, video data linguistic analysis, speech recognition module, NURBS (Non-Uniform Rational B-Splines) interpolation and geometry and gesture markup language.

Originally, for the recording and definition of handshape and gestures, many candidate methods were considered. Image processing approaches based on edge detection, motion tracking and semantic pattern recognition dominated this field. A more prominent method included haptic devices to record the signers hand movement and gestures. A very interesting, but also costly and complicated to tune, approach. Although both approaches have very good results it was agreed that, given that the HamNoSys notation commands would provide acceptable quality, they would not be implemented. In any case, semantic notation is a far more flexible and reusable solution.

3.2 Scripted signing

For the content designer to interact with a VC, a scripting language is required. In our implementation, we chose the STEP language (Scripting Technology for Embodied Persona) [8] as the intermediate level between the end user and the virtual actor. A major advantage of scripting languages such as STEP is that one can separate the description of the individual gestures and signs from the definition of the geometry and hierarchy of the VC; as a result, one may alter the definition of any action, without the need to re-model the virtual actor. The VC utilized here is compliant with the h-anim standard, so one can use any of the readily available or model a new one.

Scripted animation is an interchangeable and extensible alternative of animation based on motion capture techniques. One can think of the relation between these two approaches similarly to the one between synthetic animation and video-based instructions: motion capture can be extremely detailed with respect to the amount and depth of information, but is difficult to adjust or adapt when produced and typically requires huge amounts of storage space and transmission capacity to deliver. On the other hand, scripted animation usually requires manual intervention to compile and thus is minimal and abstract in the way it represents the various actions of the avatar. As a result, such scripts require a few hundred characters to describe and can be reused to produce different instances of similar shape [6]. This is illustrated in the code snippet in Figure 2, which illustrates the required transformations for the right hand to assume the "d"-handshape. As is easily demonstrated, 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 merely introduce the extra components into it.

In the SYNENNOESE project, a syntactic parser decodes the structural patterns of written Greek and matches them into their equivalents in GSL [2]. These are fed into an automated system that decodes HamNoSys notation sequences for each lemma; this system essentially transforms single or combined HamNoSys symbols to sequences of scripted commands. A typical HamNoSys notation sequence consists of symbols describing the starting point configuration of a sign and the action that the signing consists of. Symbols describing the initial configuration refer to the handshape that is used during the sign and the starting position and orientation of the hand that performs the sign; if the other hand takes part in the sign, as is the case in the GSL version of "doctor", it is the relative position of the two hands that matters, for example "the main hand touches the elbow of the secondary arm". Other information includes symmetry, if both hands follow the same movement pattern and any non-manual components. The VC shown here is "yt", by Matthew T. Beitler, available at http://www.cis.upenn.edu/~beitler. A demonstration with limited vocabulary and some phrase examples can be found online at http://www.image.ece.ntua.gr/~gcari/gslv.

```
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),
```

])

Figure 2. STEP code for a handshape

Figure 3. The HamNoSys sequence for the GSL version for "donkey"

Figure 3 shows the HamNoSys sequence for the particular sign, shown on the top of the page of the user interface. The first symbol here indicates that both hands perform the same movement, starting from symmetrical initial locations with respect to the signer's torso. The second symbol indicates the handshape, which here is an open palm, referred to as the "d"-handshape in GSL, while the next shows palm orientation. The following symbols handle the starting position of the palm, which here almost touches the temple of the signer's head. Symbols contained in parentheses describe composite movements, while the last character forces the signer to repeat the described movement.





Figure 5. The GSL version of

Figure 4. The GSL version of "child"



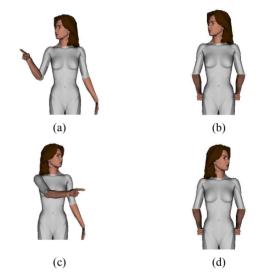


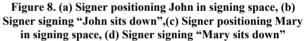
Figure 6. The GSL version of Figure 7. The frontal view of "day" "two days"

Figure 4 shows the VC signing the GSL version of "child", while Figure 5 shows an instance for the sign for "children". The design of the automated script production system enables us to use the description of the former sign to construct the definition of its plural form. In this case, the plural form is shown by repeating the same downward hand movement, while moving the hand slightly to the signer's right; direction is indicated by the symbol preceding the parenthesis, while its content describes this secondary movement. As a result, it is only necessary for the parser to indicate the particular modification of the initial sign required to produce the plural form of the lemma. In GSL, these forms are limited, thus enabling us to come up with efficient production rules, such as the one described above. Another possibility is to change the handshape for a sign, especially when the signer wants to indicate a particular quantity or number. Figure 6 shows the VC signing the GSL version of "day", while Figure 7 shows the GSL version of "two days": the difference here is that in the latter case the VC uses a two-finger handshape, instead of the straight-index finger handshape, to perform the same movement, starting from the same initial position. This difference is more evident in Figure 7, which shows the VC in a frontal view. Despite the default tilted view being the one of choice from the part of the users, the ability to show frontal and side view of the sign is crucial in learning environments, since it caters for displaying the differences between similar signs and bring out the spatial characteristics of the sign [10].

3.3 Eye gazing and technical issues

Eye gazing and head direction in all sign languages, GSL not being an exception, is far more important than in verbal languages. Where encountered it has a special grammatical meaning and frequently although the gesture of some signs are facsimile, eye gazing or head direction is the feature that differentiates them. When signing motion verbs the head and eye movement accompanies the hand movement. The most characteristic example of eye gazing use is perhaps its usage in narrations. By default the signer faces his interlocutor. When a third person is included in the plot of the narration, the signer's gaze is slightly turned so as to reference the presence of the third person. Again, when the course of events requires the signer refers to the third party by directing his head towards the position in space where he previously placed the person. Using the same pattern more than one person, that is not present, can participate in the narration by placing multiple instances of persons in different positions in space.





A nice and insightful example is the case of a narration including two persons, e.g. John and Mary, apart from the interlocutor and the signer, where John performs action "a" and Mary performs action "b" (see Figure 8). Firstly, the signer positions John and Mary in spatially specific places in the signing space, and during the signing of action "a" his gaze is directed to the correspondent position of John and respectively for action "b". In cases of an assumption the signer faces an up-right direction thus revealing the abstract meaning of his sentence. Denial is also denoted by head and eye movement. In this case not only the head, which is leaned slightly backwards, participates in the impression of the syntactic content but also the eyes gazing direction is targeted up rather than their default position, which is front.

A signer can also state time in a sentence using eye gaze. Thus signing exactly the same gestures but differentiating the eye gaze of the signers assigns a completely diverse meaning to the sentence from a temporal point of view. Finally when the signer needs to distinguish the person responsible for the action described. In the case of the 2nd person the signers gaze is attached to the direction of the signing hand. On the contrary when the active person is not participating in the discussion the head direction is different than that of the hand signing the action or verb.

The issue of eye gazing towards the signers palm, which could be considered as an end effector of a chain of links in robotics, during the animation of the sign was tackled as a combination of rotating vectors about an arbitrary axis and standard forward kinematics [12]. Thus, given the rotation axis and the relevant angle at the shoulder and elbow joints, we 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 [7]:

- since the signer is looking straight ahead, $N_{\text{current}} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$
- $N_{target} = \frac{P_{wrist} P_{skullbase}}{|P_{wrist} P_{skullbase}|}$
- $Axis = N_{target} \times N_{current}$

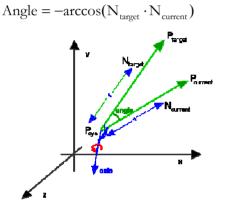


Figure 9. Overview of vectors and angles used in eye gazing [7]

3.4 Facial Expression

Facial expressions in sign languages are very important because they express the grammar. They are referred to as non-manual grammatical markers, non-manual behaviors and/or non-manual signals. Facial expressions are rule-governed. Facial expressions for questions that require YES/NO answers are different from facial expressions for wh-question words, e.g. who, why, when, where, etc. If you change the facial expressions you could convey an entirely different message. They are used in conjunction with word signs and fingerspelling to communicate specific vocabulary, questions, intensity, and subtleties of meaning. In sign language, facial expression including the raising or lowering of the eyebrows while signing is an integral part of communicating [4]. These actions help give meaning to what is being signed, much like vocal tones and inflections give meaning to spoken words. Some important functions of prosody are:

- to delimit syntactic and semantic units within sentences
- to indicate what information is focused in a sentence
- to convey pragmatic notions like illocutionary force
- to convey nuances of meaning.



Figure 10. Facial expressions of "wh-question" (a) and emphasis (b)

(a)

In this context the absence of facial expressions would be the spoken equivalent of not asking a question at all. Another usage of these expressions is the conveyance of feeling, interest, or focus. This way the signer can exhibit a range of emotion from sadness to excitement. Facial expressions can transmit interest, enthusiasm, or cheerfulness, depending on the subject matter of the presentation. For example, when a joke or humorous story is being interpreted, it is certainly acceptable - and even expected that the interpreter smile. A deadpan, expressionless face is usually considered undesirable as this can be received by the deaf as indication that the interpreter is bored.

Such expressions are not included in HamNoSys resulting in a gap in the complete notational coverage of GSL. Manual intervention was the only possible solution as to reduce the effect of this omission. Injecting information relevant to facial expression, in between the words of a phrase or in parallel to the execution of a word sign, ensures the correct and complete semantic and grammatical synthesis of GSL.

4. IMPLICATIONS AND EXTENSIBILITY **OF THE EDUCATIONAL PLATFORM**

As an educational tool above all, the SYNENNOESE platform offers a user-friendly environment for young deaf pupils aged 6 to 9, so they can have visual translation of words and phrases. The signed feedback acts as a motivating tool for spelling Greek words and structuring sentences correctly, as well for evaluating one's performance. For deaf young students as a group with special needs, the platform draws some of the accessibility barriers, and the possibility of home use even makes it accessible to family, thus encouraging communication in GSL, but also access to the majority (Greek) language. New written texts can be launched, so SYNENNOESE may receive unlimited educational content besides primary school grammar units. On the other hand, unlimited school units, such as the increasing special units with individual deaf students in remote areas can link with one another via the SYNENNOESE platform.

Moreover, text-to-sign translation can be extended and applied to different environments such as Greek language teaching to deaf students of higher grades, GSL teaching for hearing students, Greek for specific purposes such as to adult literacy classes for the Deaf etc. In this context, more domains of GSL grammar can be described and decoded, making the output closer to natural signed utterances as our analysis proceeds. This is a challenge not only for theoretical research, but also for computer science and applied linguistic research.

Furthermore, a database with the bulk of GSL utterances, described as to their features from the phonological up to the pragmatic level will be the major outcome of the whole project. In this way the representation of GSL structures can be matched to the equivalent ones of written Greek, and it will be a challenge to be able to compare directly the grammars of the two languages. In much the same way structures of GSL will easily be compared with counterparts from ASL or BSL [3] for research across signed languages.

4.1 PROBLEMS AND LIMITATIONS

The main limitations of the study are described below. These are divided into linguistic, educational and technical ones. Most of the limitations are typical to sign animation projects, and they were expected before the beginning of the project.

Regarding the linguistic and educational aspects of the project, one of the major issues that needs to be addressed is the fact that in some areas of the language there are no standardized signs, so there may be some theoretical objections as to the use of particular entries. However, a platform such as the one described allows for multiple translations and does not have any limitations as to the size of files, which was the case, for example in previous DVD-based GSL dictionaries, containing video entries. Moreover, the platform will be open to updates through the script authoring process.

Besides this, the data available in GSL, when compared with data from written Greek, for example, are dauntingly scarce. Error correction mechanisms were sought after in order to assure reliability of results. Such back-up mechanisms are the use of approved dictionaries, the consultancy of Pedagogical Institute and the feedback from the Deaf Community, along with the continuing data from GSL linguistic research.

The ultimate challenge, as in all similar projects, remains the automatic translation of the language. It is still too difficult to produce acceptable sentences in the automatic translation of any language at the moment, even more so a minor, less researched language with no written tradition such as GSL. Realistically the teams involved in the SYNENNOESE project can expect as an optimum result the successful use of automatic translation mechanisms in GSL only in a restricted, sub-language oriented environment with predetermined semantic and syntactic characteristics.

5. CONCLUSIONS

In this paper we have described the underlying design principles and implementation of a web-based virtual signer software component, utilizing language resources suitable for young pupils. This component uses standard linguistic and virtual character technologies to provide semantic and syntactic information from written text and encode it with reusable and extensible sign notation representations. These representations are readable by the VC platform, making them suitable for teaching GSL and providing signed summaries of documents.

6. REFERENCES

- [1] Bellugi, U. and Fischer, S., A comparison of Sign language and spoken language, *Cognition*, 1, 173-200, 1972
- [2] Boutsis S., Prokopidis P., Giouli V. & Piperidis S., Robust Parser for Unrestricted Greek Text. Proc. of the 2nd Language Resources and Evaluation Conference, 467-474, 2000.
- [3] Brien D. and Brennan, M, Dictionary of British Sign Language / English. Faber and Faber, Boston, 1992.
- [4] Brock Craft, D.H., Rosalee W., et al., An Approach to Modeling Facial Expressions Used in American Sign Language, in DePaul CTI Research, 2000, Chicago, IL.
- [5] Chittaro, L. and Serra, M., A Brief Introduction to Web3D Technologies in Education: Motivations, Issues, Opportunities, *in Proc. of LET-WEB3D 2004*, Udine, Italy.
- [6] Foulds, R., Biomechanical and Perceptual Constraints on the Bandwidth Requirements of Sign Language, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, March 2004, pp. 65–72.
- [7] http://www.euclideanspace.com/maths/algebra/vectors/lookat /index.htm
- [8] Huang, Z., Eliens, A., and Visser, C., STEP: A Scripting Language for Embodied Agents, *Proceedings of the Workshop on Lifelike Animated Agents*, 2002.
- [9] ISO/IEC 19774 Humanoid Animation (H-Anim), http://www.h-anim.org
- [10] Kennaway, R., Experience with, and Requirements for, a Gesture Description Language for Synthetic Animation. 5th International Workshop on Gesture and Sign Language based Human-Computer Interaction, Genova, 2003.
- [11] Kyle, J. G. and Woll, B., Sign Language: the study of deaf people and their language, Cambridge University Press, 1985.
- [12] Lengyel, E., Mathematics for 3D Game Programming and Computer Graphics, Charles River Media, Hingham, USA
- [13] Liddell, S., *American Sign Language Syntax*, The Hague: Mouton, 1980.
- [14] Padden, C., The Relationship between Space and Grammar in American Sign Language Verb Morphology, *Theoretical Issues in Sign Language Research*, Gallaudet University Press, 1989.
- [15] Prillwitz et al, HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide, 1989.
- [16] Stokoe, W. and Kuschel, R., For Sign Language Research, Linstock Press, 1978.
- [17] Wilcox, S., Scheibmann, J., Wood, D., Cokely, D. & Stokoe, W., Multimedia dictionary of American Sign Language, *Proceedings of ASSETS Conference*, pp. 9-16, 1994.