DICTA-SIGN Publishable Summary 2011-2012

"Sign Language Recognition, Generation and Modelling with Application in Deaf Communication"

URL: http://www.dictasign.eu/

DICTA-SIGN researches ways to enable communication between Deaf individuals through the development of human-computer interfaces (HCI) for Deaf users, by means of Sign Language.

TABLE OF CONTENTS

PROJECT DESCRIPTION

DICTA-SIGN is a three-year EU-funded research project that aims to make online communications more accessible to Deaf sign language users.

The development of Web 2.0 technologies has made the WWW a place where people constantly interact with each other, by posting information (e.g. blogs, discussion forums), modifying and enhancing other people's contributions (e.g. Wikipedia), and sharing information (e.g., Facebook, social news sites). Unfortunately, these technologies are not easily accessible to sign language users, because they require the use of written language.

Can't sign language videos fulfill the same role as written text in these new technologies? In a word, no. Videos have two problems: Firstly, they are not anonymous – anyone making a contribution can be recognized from the video, which holds many people back who would otherwise be eager to contribute. Secondly, people cannot easily edit and add to a video that someone else has produced, so a Wikipedia-like web site in sign language is not possible based on video.

DICTA-SIGN's goal is to develop the necessary technologies that make Web 2.0 interactions in sign language possible: Users sign to a webcam using a dictation style. The computer recognises the signed phrases, converts them into an internal representation of sign language, and then has an animated avatar sign them back to the users. Content on the Web is then contributed and disseminated via the signing avatars. Moreover, the internal representation also allows us to incorporate simple sign language-to-sign language translation services, based on sign correspondence in different sign languages.

In this way, DICTA-SIGN aims to solve both of the problems that sign language videos have. The avatar is anonymous, and its uniform signing style guarantees that contributions can be easily altered and expanded upon by any sign language user.

RESEARCH & DEVELOPMENT AREAS OF INTEREST

DICTA-SIGN deals with four Sign Languages: British Sign Language (BSL), German Sign Language (DGS), Greek Sign Language (GSL) and French Sign Language (LSF).

The project involves research from several scientific domains in order to develop technologies for sign recognition and generation, exploiting significant knowledge of the structure, grammar and lexicon of the project Sign Languages, the so called linguistic knowledge and resources of a language. Sign Language linguistic knowledge can be derived exclusively by appropriate processing of Sign Language video corpora, linked to grammars and lexicons.

To serve its goals, DICTA-SIGN combines linguistic knowledge with computer vision for image and video analysis that serves to achieve continuous sign recognition as presented in sign language videos, and with computer graphics for realistic signing animation by means of a virtual signer (avatar).

SUMMARY OF ACTIVITIES

The third and final year of the DICTA-SIGN project has culminated in prototype applications for sign look up, translation and wiki abilities. This has been possible via research advances and continued annotation of the project's multilingual sign language resources. In parallel, the project's outcomes have been introduced to Deaf end users for extensive evaluation, while a number of dissemination activities have promoted the project's progresses to the international scientific community and the European Deaf Community.

With the arrival of the Kinect at the end of the second year, the DICTA-SIGN team responded well and worked on extracting 3D features from this device. Sections of the collected corpus were also tracked in 3D to complement the Kinect data (Figure 1). This was done using a novel scene flow approach on the wide-base stereo data.

Figure 1: 3D tracking of corpus data

From this tracking, features are extracted which relate the information to linguistics. An example of these features is shown in Figure 2. This uses the Kinect skeleton to extract which part of the signing space is being used. Other types of features extracted are those relating to motion and handshapes.

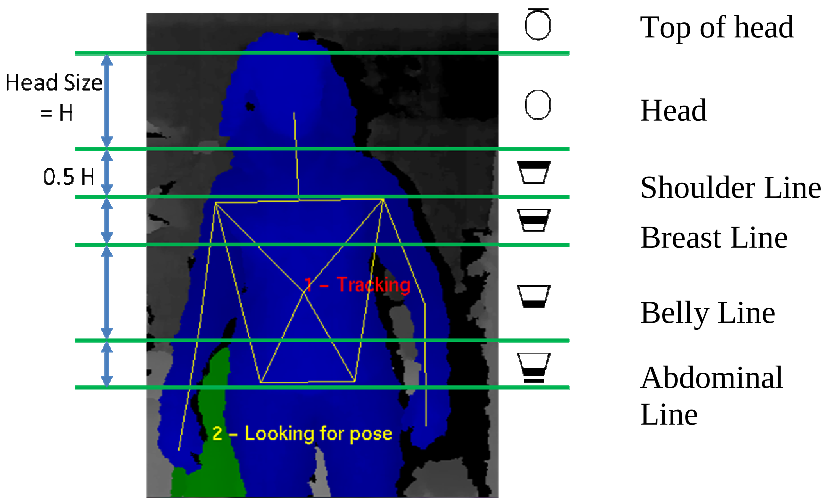

Figure 2 Example features extracted from 3D tracking

Also being tracked are the facial features of the signer (Figure 3). These are being used to learn geometric feature combinations which relate to Non-Manual-Features annotated in the corpus. The face tracking and feature extraction in Fig. 3 is based on Active Appearance Models, applied both globally onto the whole face as well as locally on the eye, eyebrows and mouth regions of a signer's face.

Figure 3 Tracking of the facial features

Finally, *video preprocessing, hand & face tracking* as well as *movement and handshape feature extraction* on parts of the *continuous Dicta-Sign SL corpus* has been performed. The hand tracking is based on skin color modeling with Gaussian Mixture Models, by also taking into account possible occlusions. Extraction of handshape features on the continuous corpus has been based on the developed approach for Shape-Appearance Affine invariant handshape modelling and feature extraction (Figure 4).

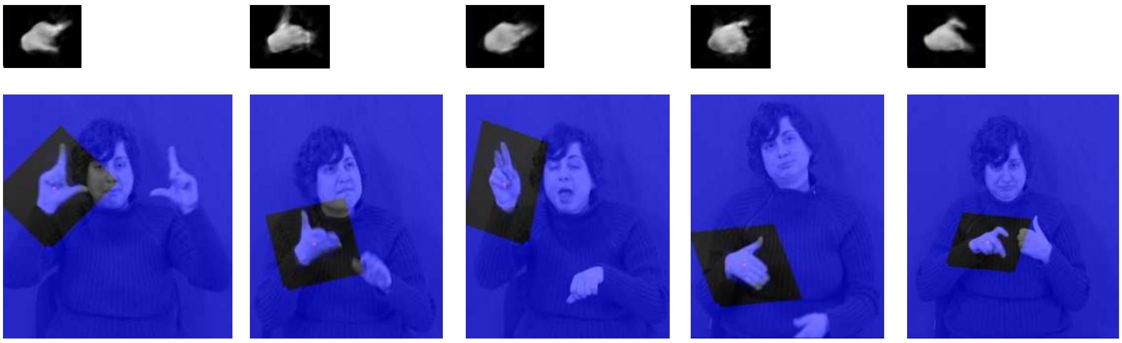

Figure 4: Regularised Shape-Appearance model fitting on continuous SL data: handshape feature extraction

Concerning Sign Language Recognition, progress on the following has been achieved: (1) *Continuous Modeling - Phonetic Sub-Units (SU)s*: this direction produces sub-units that are more meaningful than the data-driven ones. Such sub-units are novel in this field and provide a significant advantage for both interdisciplinary research as well as for other tasks of continuous recognition. (2) *Co-articulation and Relocation of Signs* for continuous sign recognition by accounting for mismatches between the citation form of the signs and the actual articulation. (3) *Models of Multiple Signers* for the Dicta-Sign GSL Lemmas Corpus. (4) *Fusion of multiple cues/modalities* in sub-unit level and continuous data for sign recognition, where we exploit our previous developed approach of dynamic-static sub-units for movement-position+handshape features. (5) *Facial Features Classification and correspondence with extreme facial event detection*.

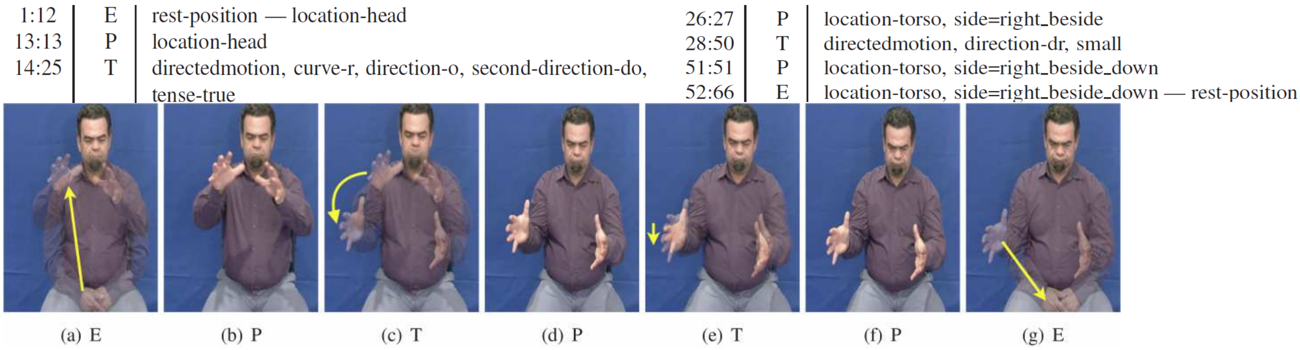

Figure 5: Recognition process for GSL sign PILE

As regards *Dynamic-Static Integration of Movement-Handshape for Sign Language Recognition*, the integration of movement-position (MP) and handshape (HS) cues have been explored in order to combine the data-driven subunits (SUs) modeling the dynamic-static notion for MP and the affine shape-appearance SUs for handshape configurations, leading to a new dynamic-static integration of manual cues. In Continuous Sign Recognition another major issue that has to be dealt with is the *Coarticulation and Relocation of Signs*: both phenomena have been examined and a generic network which allows the insertion of epenthesis transitions and/or postures between signs was employed. Relocation for signs that can be articulated in different positions has also been addressed with related transformations on the lexicon that allow one sign to be articulated in multiple different positions. An example of the decoding output in continuous signing of Kinect tracking data is illustrated in Figure 6.

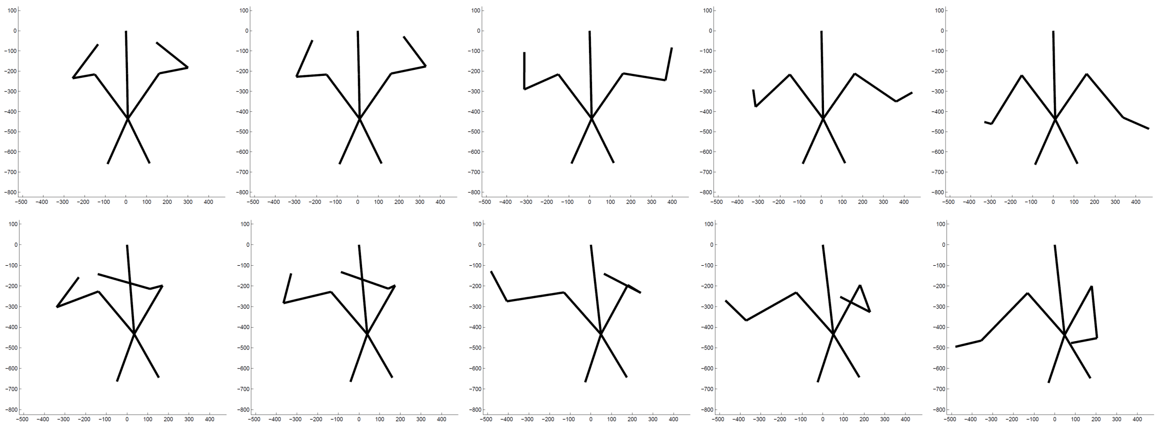

Figure 6: Demonstration of coarticulation-relocation of signs for the Continuous Sign Recognition framework

Sign language linguistic models developed by the project at the lexical level enable inferences to be made about features that vary during sign language performance. Examples include the locations at which signs are performed, and the movements linked to directional verbs that identify pronouns. Work has been done to build on partial recognition of signs that identifies a subset of linguistic features (expressed in HamNoSys). Fully inflected sign sequences can be inferred for a class of sign language phrases.

Beside the lexical level, and via the multilevel model Azalee, we also studied linguistic structures, based on the DICTA-SIGN corpus data. Three languages were included in the study, which constitutes the first cross-linguistic formalisation of structures relevant to all levels of language. We have studied six structures: alternative options listing, enumeration, some cases of naming, title/topic announcement, neutral questions and quantifier structures. It was interesting to find that some of these structures were expresses with language-specific mechanisms n a greater extend than others, whereas some invariants work for all languages of the study. A number of the latter have been implemented to integrate their synthesis in the wiki.

Work on Sign Synthesis and Animation during the past year has created a DICTA-SIGN avatar, Françoise, that can be used in more than one animation framework, thus allowing interoperability between systems.

The avatar, and animations developed for her, can also be exported to industry-standard packages such as Maya and 3ds Max.

In addition to the Java-based renderer used by the JASigning system, a new renderer has been developed using WebGL, the emerging standard for 3D animation in browsers using HTML5.

Enhancements to the JASigning system have continued, especially to provide amore accurate animation of handshape changes. This work follows from a study of the DICTA-SIGN corpus material and is reported in a paper presented to the SLTAT workshop in October 2011. Other JASigning changes have supported the development of the project prototypes and demonstrators.

Figure 7: The DICTA-SIGN avatar Françoise

A major focus of the third year of the project has been systems integration, linking recognition technology, via linguistic processing, to synthesis and animation. The search-by-example interface to the DICTA-SIGN lexical database has been enhanced to produce a Sign look-up tool that provides users with a simple sign-level translation tool for exploring corresponding signs in the four project sign languages.

The sign wiki prototype was identified as the most useful demonstration of DICTA-SIGN technologies for Deaf people.

PROJECT APPLICATIONS & USER INVOLVEMENT

Project Applications

During the third year of project life cycle, DICTA-SIGN has successfully integrated technologies and resources developed within the project. These take the form of proof-of-concept end user application prototypes; a *Sign look-up tool* and a *Sign Language Wiki,* which best demonstrate the potential of sign language incorporation in web applications.

Sign look-up tool

Natural sign languages differ from one another, just like spoken languages. While some signs are iconic, and are relatively similar in all the languages used by the project, others are very different. The Sign look-up tool enables a Deaf user to perform a sign and see the corresponding sign in the four languages used by the project.

The system uses a Microsoft Kinect™ device to recognise signs:

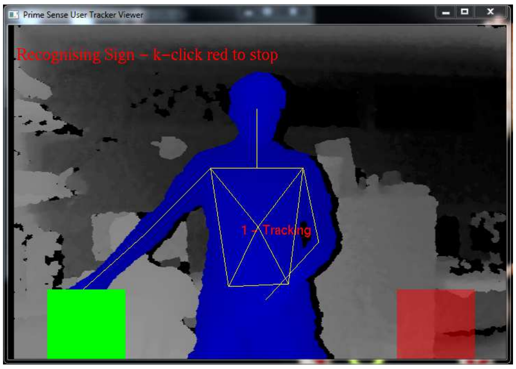

Figure 8: Microsoft Kinect input

Signs are matched using a robust, discriminative, classification architecture based on sequential pattern trees. This novel approach results in a user-independent system which can work in varied environments. The Sign look-up tool plays back the closest matches to the sign via an avatar. The user can then see the corresponding signs in all four languages.

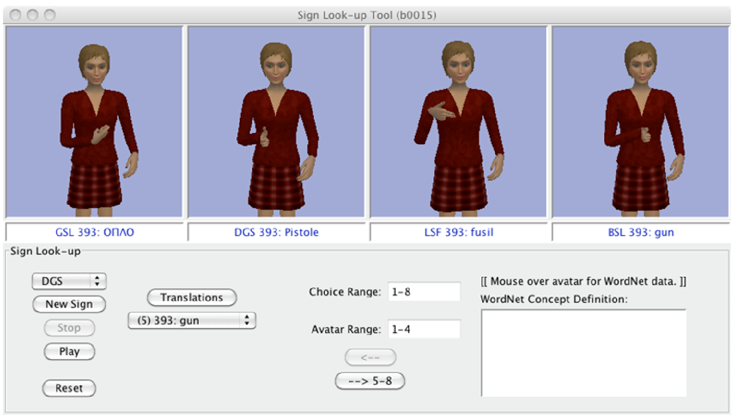

Figure 9: Sign look-up tool results

A practical use for this tool is where a sign language user sees a sign in an unfamiliar language on a video, or when travelling to another country. They can perform this sign to the look-up tool and if the sign is recognisable to the system, they can be shown a version in their own language.

Sign language Wiki (Sign Wiki)

A major requirement of contemporary Web 2.0 applications is that user contributions are editable by an entire community. The oldest, and most popular, application of this type is a Wiki, where any contribution can be edited and refined, anonymously if so wished, by someone else. As the success of Wikipedia and related sites show, this type of community collaboration results in a rapid amassing of knowledge.

There is no doubt that sign language users could benefit similarly from collaborative editing. A server is developed providing the same service as a traditional Wiki, but using sign language. Instead of using text as the output medium, a signing avatar presents information:

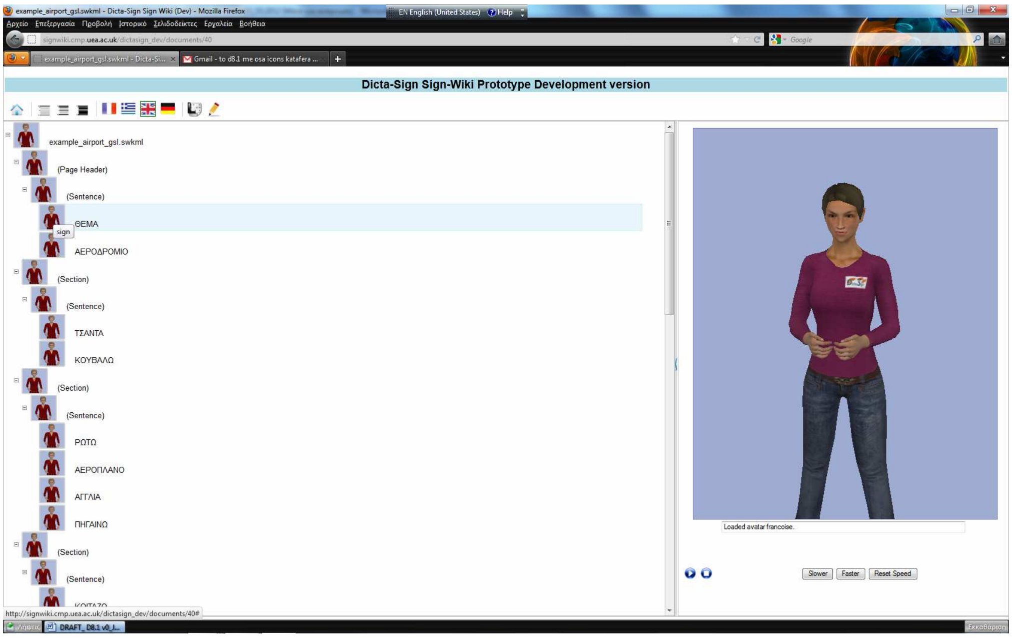

Figure 10: Standard Sign Wiki page format / Viewing mode

The use of an avatar preserves the anonymity of the user, and facilitates modification and reuse of information present on the site.

The system acts as a dictaphone using sign, providing recording, playback, and editing. A user can put information onto the server using sign language by means of the Microsoft Kinect device. The system then matches the user's signs against a stored dictionary, and the matched signs are used to generate the movements of the signing avatar. The user then reviews the recognised signs:

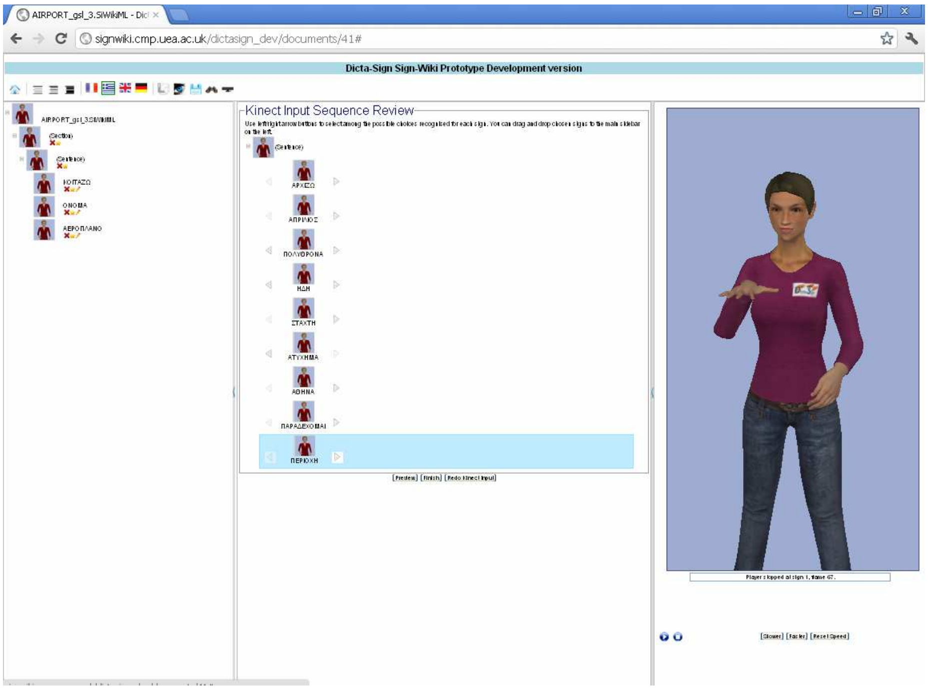

Figure 11: DICTA-SIGN Sign-Wiki input reviewing environment

If the system interprets the sign language sentence incorrectly, the user can correct it by repeating only the erroneous part.

Editing of individual features is also supported to change the initial posture of a sign:

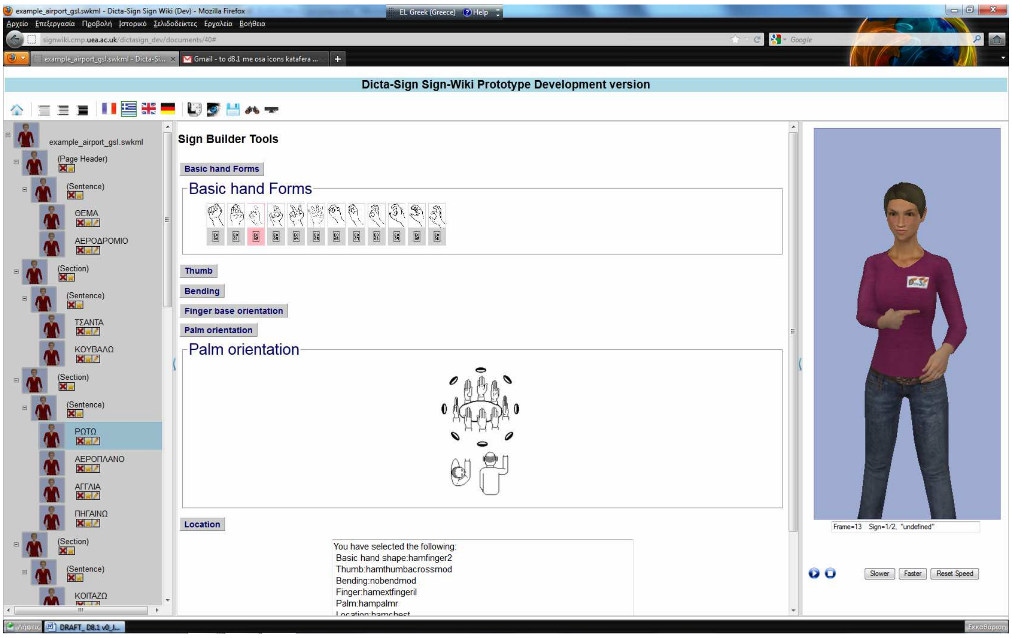

Figure 12: DICTA-SIGN Sign-Wiki sign building environment

If a Kinect device is not available, sequences can be created by selecting signs from the lexicon, using spoken language synonyms to expand the range of choices:

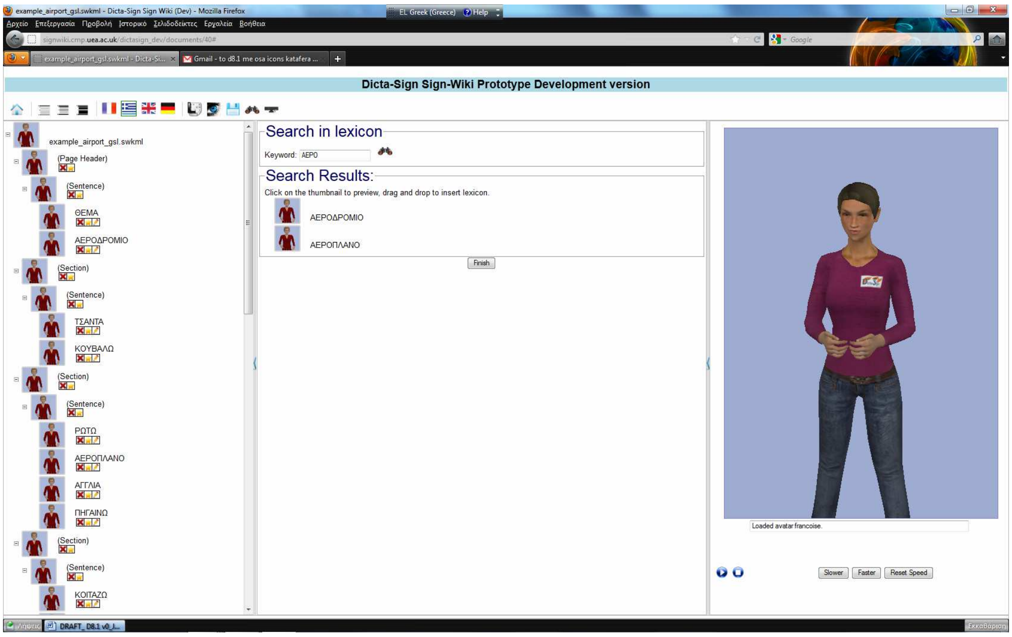

Figure 13: DICTA-SIGN Sign-Wiki lexicon search results

User Involvement & Evaluation

It is a major concern of DICTA-SIGN that project results reach the Deaf Communities of the partner countries and also that Deaf users are actively involved in evaluation of all project's applications. During its third year, the project has been opened to Deaf users for extensive evaluation, while Deaf working group members have undertaken wide dissemination to the international Deaf community.

DISSEMINATION and FUTURE EXPLOITATION PROSPECTS

Dissemination Strategy

DICTA-SIGN web site

The DICTA-SIGN website (www.dictasign.eu) is the project's main communication tool. It contains multilingual (written form of oral language and SL videos) material that reflects the project's aims, research progress and scientific impact. This is the place where all information related to DICTA-SIGN is stored and made accessible to the Internet sharing community.

The home page provides project descriptions in the eight languages of the project (English, French, German, Greek, BSL, DGS, GSL and LSF).

As the project progresses, the website provides information and allows public access to the following topics:

- Project overview
- Consortium data
- Scientific publications of project research groups
- Project applications

- Project newsletter
- News and Events
- Publicly available project deliverables

During 2011, the DICTA-SIGN website has presented considerable visibility with 6.064 distinct visitors and 19.575 visits.

Dissemination to the scientific community and the industry

Dissemination to scientific community has continued during the third project year in the manner already adopted in the two previous years. This is based on organisation of significant scientific events, bilateral exchange of information with major scientific institutions, as well as communication of project achievements in conferences and through publications (http://www.dictasign.eu/Main/Publications) .

As part of this, DICTA-SIGN partners took a leading role in organising two international workshops on Sign Language Translation and Avatar Technology. The first SLTAT Workshop was held in Berlin sponsored by the German Federal Ministry of Labour and Social Affairs (BMAS) in January 2011 (http://embots.dfki.de/SLTAT11-Berlin/). It featured several presentations by DICTA-SIGN partners. The second SLTAT workshop was held in conjunction with ASSETS 2011, the ACM SIGACCESS Conference on Computers and Accessibility, at the University of Dundee, Scotland, in October 2011 (http://embots.dfki.de/SLTAT/). Four of the papers and posters at this workshop reported DICTA-SIGN work.

In addition the DICTA-SIGN coordinator organised GW2011: The 9th International Gesture Workshop "Gesture in Embodied Communication and Human-Computer Interaction" http://access.uoa.gr/gw2011, in May 2011, in Athens, Greece.

Promotion of the project's innovative technologies also continued via a framework of national and international conferences, exhibitions and scientific events most attractive to the industry.

COLLABORATION

Collaboration between consortium members

The consortium was formed on the basis of a long tradition of academic contacts among most of the partners. The working groups in DICTA-SIGN are formed on the basis of laboratory expertise and specific project implementation needs.
During the third year of the project, its 6 working groups, entailing:

- The Sign Language Recognition Group,
- The Sign Synthesis and Animation Group,
- The Grammar Modelling Group,
- The Annotation Tools Group and
- The Parallel Corpora Group

are fully active and their synergies have led to successful integration of all foreseen project prototypes.

Clustering activities

DICTA-SIGN partners participate in various activities of knowledge sharing within their domains of expertise. In the framework of GW-2011, exchange of scientific advances and achieved goals in the areas of sign language corpus research and sign language technologies included representation of the SignSpeak (www.signspeak.eu) project, aiming at fruitful clustering activity between the two projects. Clustering activity extends also beyond the life cycle of both projects in the framework of the 5th Workshop on the Representation and Processing of Sign Languages (LREC-2012).

Events organised by DICTA-SIGN partners during 2011

- **First International Workshop on Sign Language Translation and Avatar Technology (SLTAT)**. Berlin, 10-11 January 2011: http://embots.dfki.de/SLTAT11-Berlin/
- **Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)**. 23 October 2011. University of Dundee, UK: http://embots.dfki.de/SLTAT/
- **GW2011: The 9th International Gesture Workshop "Gesture in Embodied Communication and Human-Computer Interaction**", May 25-27, 2011, Athens, Greece: http://access.uoa.gr/gw2011

Forthcoming events organised by DICTA-SIGN partners

- **The 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon.** Satellite Workshop to the eighth International Conference on Language Resources and Evaluation (LREC-2012), 27 May 2012, Istanbul, Turkey: http://www.sign-lang.uni-hamburg.de/lrec2012/cfp.html

USEFUL LINKS

DICTA-SIGN consortium and contact persons

Athena RC, Institute Language and Speech Processing — Greece

Dr Eleni Efthimiou elenie_e(_@_)ilsp.gr
Dr Stavroula-Evita Fotinea evita(_@_)ilsp.gr

University of Hamburg, Institute of German Sign Language and Communication of the Deaf — Germany

Thomas Hanke thomas.hanke(_@_)sign-lang.uni-hamburg.de

University of East Anglia, School of Computing Sciences — United Kingdom

Pr John Glauert j.glauert(_@_)uea.ac.uk

University of Surrey, Centre for Vision, Speech and Signal Processing — United Kingdom

Pr Richard Bowden r.bowden(_@_)surrey.ac.uk
Pr Josef Kittler j.kittler(_@_)surrey.ac.uk

LIMSI-CNRS, Sign Language Processing team — France

Dr Annelies Braffort annelies.braffort(_@_)limsi.fr
Annick Choisier annick.choisier(_@_)limsi.fr

Université Paul Sabatier, IRIT — France

Pr Patrice Dalle dalle(_@_)irit.fr
Dr Christophe Collet collet(_@_)irit.fr

National Technical University of Athens — Greece

Pr Petros Maragos maragos(_@_)cs.ntua.gr
Vassilis Pitsikalis vpitsik(_@_)cs.ntua.gr
Stavros Theodorakis sth(_@_)atrion.gr

websourd — France

Dr Jeremie Segouat jeremie.segouat(_@_)websourd.org
Dr François Lefebvre-Albaret françois.lefebvre-albaret(_@_)websourd.org
Julia Pelhate julia.pelhate(_@_)websourd.org