Terminology Enrichment through Crowd Sourcing at PYLES Platform

Eleni Efthimiou[1], Stavroula-Evita Fotinea[1], Panos Kakoulidis[2], Theodoros Goulas[1]

[1]Institute for Language and Speech Processing (ILSP)/R.C. "Athena", [2]Innoetics
[1/2]Artemidos 6 & Epidavrou, Maroussi, 15125 Athens, Greece
E-mail: {eleni_e, evita, tgoulas}@ilsp.gr, pk88gr@hotmail.com

Abstract

Here we present how Greek Sign Language (GSL) has provided content to various functionalities of the Information System PYLES, a management system for on-line lessons, designed to support accessible asynchronous e-learning by addressing learning needs of students with various communication capabilities and needs, at the Technological Educational Institute of Athens (TEI-A). In order to meet native signers' student needs, the platform has incorporated descriptions in GSL of various lessons in the curriculum, as well of administrative forms and documents. However, the most useful tool for the student community has been a bilingual on-line lexicon which provides both general purpose content and domain specific terminology glossaries. The major characteristic of this tool is that it allows for content enrichment via crowd sourcing, which, hopefully, proves to provide a satisfactory solution in respect to terminology gaps often noticed one major difficulty in of deaf individuals' education.

Keywords: SL terminology glossaries, SL lexicon resources, GSL-Greek bilingual resources, deaf education services, educational platform accessibility services, terminology crowd sourcing

1. Introduction

In many occasions there has been made reference to the difficulty deaf individuals are confronted with when they need to access a written text (Chamberlain et al., 1999; Leigh & Andrews, 2017; Ghari, 2016). This phenomenon becomes more acute in the context of Higher Education, where students are constantly in contact with written material, very often even written in another language than the one of their hearing environment.

Furthermore, terminology related issues are known to be observed in all cases, where a new term or a set of new terms created within a specific language system, need to be transferred to some hosting language system.

Although this problem has been well observed and transfer procedures are described in detail in a series of ISO recommendations (ISO/R 704, 1968; ISO/R 1087, 1969) and their updates (ISO 704, 1987; ISO 1087, 1990; ISO 1087/1, 2000) since mid 90's, the whole process of terminology transfer and validation is a difficult one and very few terms end up to be widely incorporated in the linguistic reality of the active language domain users.

This paper presents how Greek Sign Language (GSL) is incorporated in the educational material provided to deaf students of the Technological Educational Institute of Athens (TEI-A) via PYLES Information System, a management system for on-line lessons, designed to support asynchronous e-learning.

Besides the use of GSL video for the presentation of administrative forms and documents, the following sections will deal with demonstrating how accessibility to educational content is supported, focusing on an on-line lexicon that encounters a general purpose bilingual dictionary for the language pair GSL-Greek and a number of terminology glossaries from the scientific domains taught at TEI-A at an open environment that allows for contributions from different categories of users in a crowd sourcing appeal, targeting to resources enrichment, following a validation protocol that ensures quality control of the newly added lexical/ terminological items.

2. PYLES Information System Architecture

The Information System PYLES is a management system for on-line lessons, designed to support accessible asynchronous e-learning, addressing learning needs of students with various communication capabilities and needs at the Technological Educational Institute of Athens (TEI-A). It, thus, exploits both up-to-date assistive technology software and content presentation in various forms.

The platform has been used as the basis for the development of an active repository of multimodal educational resources, also incorporating a terminology section that entails domain specific terminology glossaries and a general purpose bilingual dictionary for the language pair GSL-Greek.

The platform provides advanced customization options according to user needs (Fig. 1) but also a collaborative environment for the support of teaching and learning processes (Šumak et al., 2011).

The information system [1] is built on the open code platform *Open eClass*[2], a free e-learning platform that has actually been enriched with tools and functionalities which allow extended accessibility regarding both the environment and the educational content (Efthimiou et al., 2015).

The GSL terminology environment allows for the creation of different glossaries directly by their users, where GSL signers are invited to upload their suggestions for various terms under specific quality control conditions as presented in section 2.2.

[1] http://eclassamea.teiath.gr /.
[2] http://www.openeclass.org/.

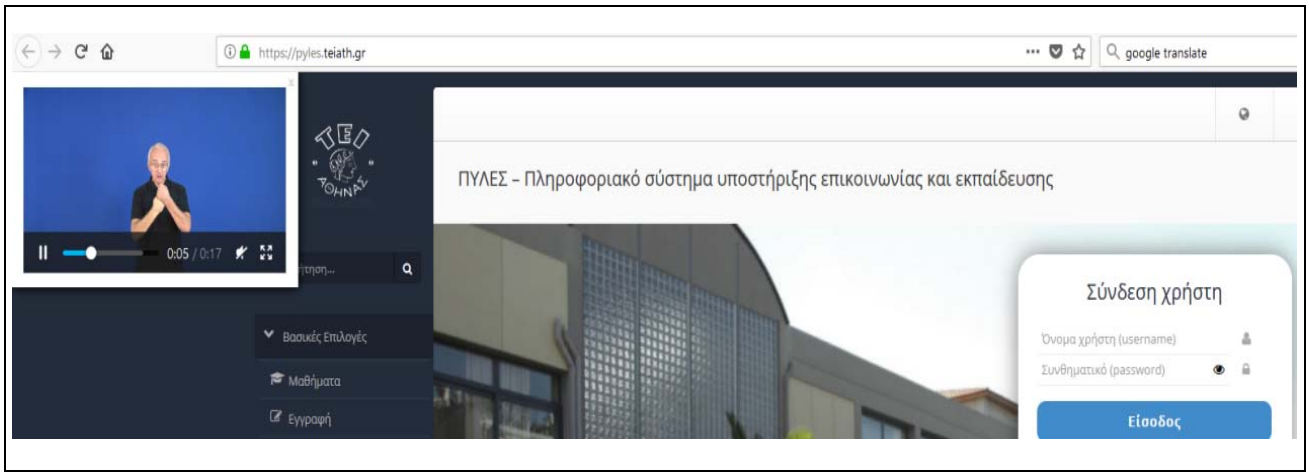

Figure 1: PYLES platform main page with activated navigation information in GSL and user login window.

Two types of terminology resources are available to all users:

a) Terms created and validated according to ISO provisions (ISO 704, 1987; ISO 1087, 1990; ISO 1087/1, 2000),
b) Terms collected via crowd sourcing.

For the search and presentation of lexical items (general purpose lexicon or terminology), the user, after logging in, may select the domain of his/her interest from the list in (Fig. 2) and then typing his/her wished item in the search box.

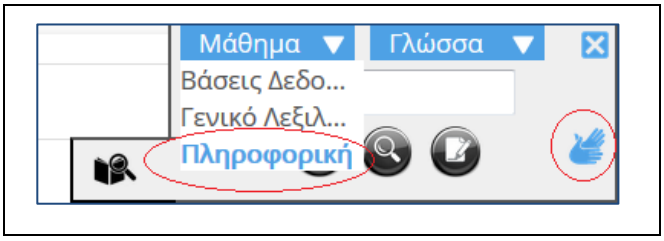

Figure 2: Selection of terminology domain from a list for the search of a term, when GSL presentation mode is activated.

Another option is to search an item from an alphabetically ordered list (Fig. 3).

When in alphabetical search environment, the user is provided also with various statistics such as "latest entrances" and "highly scored entries", along with other crowd sourcing and lemma related information such as the language domain the lemma belongs to, the date of its addition and the language(s) in which it is available.

Accessibility of content at PYLES has been supported by exploiting language based assistive technologies which involve implementation of a synthetic voice facility for the accessing of written content across platform by users with vision problems, along with incorporation of the GSL tools presented here to support GSL signers use the platform.

Activation of all accessibility facilities is subject to customization, according to specific user needs and preferences.

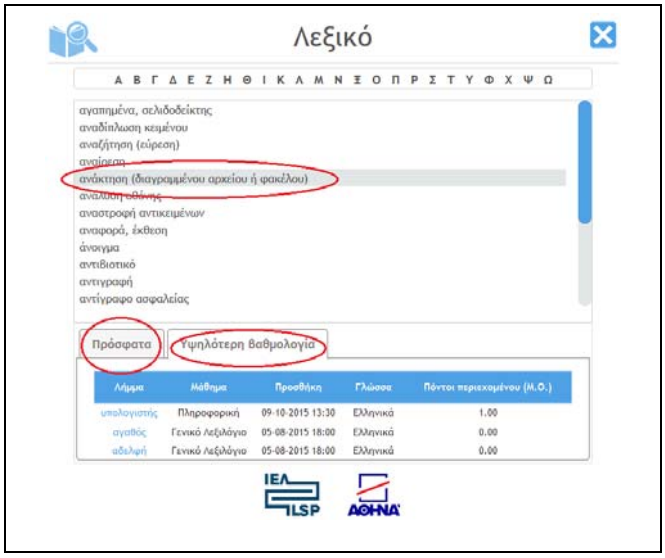

Figure 3: Alphabetical search option.

Regarding overall customization options intended to serve GSL signers' needs (Dimou et al., 2014), the platform incorporates:

- Selected lesson presentations in GSL on the basis of deaf students' preferences regarding the curriculum offers,
- An on line bilingual dictionary of general purpose lemmas for the language pair GSL-Greek,
- Online terminology glossaries which provide terminology items presentation options in GSL, Greek and English,
- Administrative form related information in GSL.

3. Terminology Resources at PYLES Platform

Conceptualization of terminology is a major factor for acquiring new knowledge and also a prerequisite for the production of new knowledge within a scientific domain (Sager, 1994). In this respect, terminology is of major importance in the framework of scientific education and vocational training.

However, when investigating the availability of terminology in various SL national systems, one discovers a critical gap that directly affects deaf students' integration to productive higher education communities.

This situation is due to a twofold problem. On the one hand, introduction of new terminology lists to a linguistic system by following international standards is a difficult and time consuming task that requires dedicated involvement of relatively big groups of experts for a considerably long period. On the other hand, deaf students/scientists usually need to acquire the terminology of their domain of expertise by accessing written texts in English. This task sets an extra burden for everyone been raised in a non English speaking environment, in addition to the generally recognized difficulty a considerable number of deaf individuals face with accessing written information.

At PYLES platform, since the major goal has been to support educational activity, it has been necessary to cope with the terminology lack problem. Thus, two options of terminology list creation have been adopted: (a) by gathering existing term collections created according to ISO recommendations, and (b) by provision for a crowd sourcing activity in respect collecting not yet fully validated (*ad-hoc*) terms, which, however, are actively in use within domain specific user communities.

3.1 Terminology Creation According to ISO Recommendations

The procedure, as foreseen by the various ISO recommendations and their updates relating to introduction of existing terminology into a new language, involves a series of steps associated with different tasks performed by working groups of experts and groups of members from the receiver language community.

More specifically, in *step 1*, a group of experts, composed by individuals with long experience of work within a given domain, examine the related terminology glossary and propose a translation for each new terminology item. This translation must be representative of the conveyed knowledge, be accepted by the technical user community, but also it must follow the rules of the receiver language as regards new lexical item formation mechanisms (Gruber, 1993; Katsoyannou & Efthimiou, 2004). These requirements make it necessary to include theoretical linguists and lexicographers to terminology translation proposing groups.

In *step 2*, the list with the experts' proposal is sent to an extensive working group consisting of representatives of user communities, in order to be checked. This second working group is obliged to return the list with markings for "acceptable", "rejected", "needs correction", "new proposal" for each terminology item undergoing the validation process.

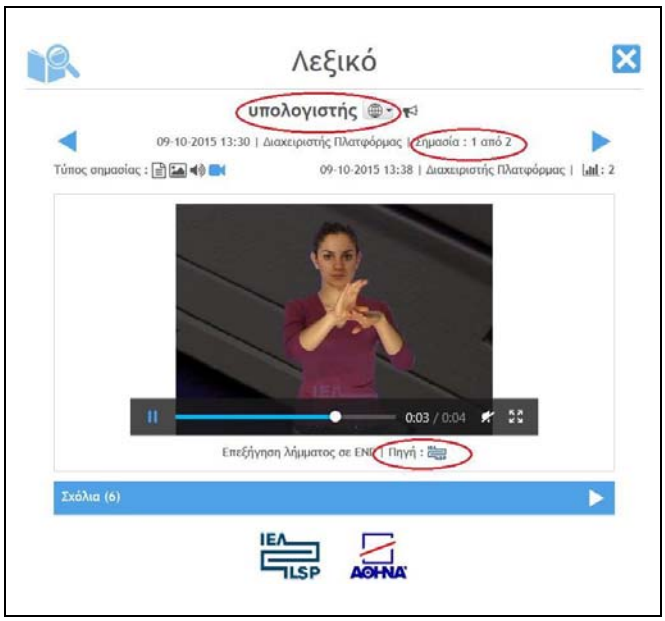

Figure 4: Term presentation window at TEI-A platform.

In *step 3*, the proposing group of experts incorporates the user community's suggestions and delivers the term list again for a second validation process.

Step 4 is the final state of a term validation process, if the user community group accepts the translation as an integral part of their language.

Steps 2 to 4 may be iteratively repeated as many times as needed in order for a term translation to be validated and accepted.

As regards terminology in the TEI-A platform, the glossary of basic computer use terms has been translated to GSL following the above described procedure, where initial suggestions have been gathered from deaf experts teaching *Introduction to Computer Use* classes for several years. Their initial video recorded glossaries have been discussed among members of native GSL signers' community for a long time and in a number of iterative sessions for steps 2 to 4 of the validation procedure before unanimous agreement was succeeded and the term collection was acquired in its final form (DIOLCOS, 2006).

But terminology items, in order to serve knowledge transfer, need also to be associated with a number of other types of information which clarify their meaning within a specific domain of use. Such information incorporates the term equivalent in other languages, a term definition and explanatory visual material like pictures, diagrams, drawings, etc. Of equal significance is reference to the source of information accompanying each term.

For the creation of the terminology presentation window at the TEI-A platform, we have taken all these parameters into account, emphasizing on the platform's educational dimension (Efthimiou & Fotinea, 2007; Efthimiou, 2008;

Fotinea & Efthimiou, 2009; Fotinea et al., 2012). All information display options relating to a specific term that are available to the user, are depicted in Figures 4 to 6. In (Fig. 4) the encircled information accompanying the term display in GSL refers to (i) the Greek equivalent term and the language button, presenting the term equivalent in more languages as i.e. English, both centered at the top of the window, (ii) the senses associated with the displayed form in the next line, and (iii) the source database of the displayed information underneath the video window, which in this case is the term "*computer*" in GSL and derives from the ILSP repository as indicated by the institute's logo.
In (Fig. 5), the definition of the term is the main displayed element. In this case, the source of the displayed information is the Greek Wikipedia, while other displayed information identifies the person who entered the definition and the time of creation of the displayed information.

Figure 5: Term definition presentation. Encircled is information about time of make and creator of entry, as well as its source.

The encircled icons at the top left side of the window indicate the options for the means of information display related to the chosen term, which are available to the user, where blue color indicates the currently chosen option.

In (Fig. 6), the display option is selected, which provides visual support material for the comprehension of the term. Furthermore, the "more languages" button (globe icon) placed next to the Greek term equivalent is also selected, displaying the term equivalent in English. The at least trilingual association of each term (GSL, Greek and English) has been considered necessary, since students at Greek universities very often need to consult bibliography written in English language. In the (Fig. 4) example the central explanatory icon source is again the Greek Wikipedia.
This terminology collection is also used as a pilot for the validation of terminology collections in other scientific domains when translated to GSL, since crowd sourced terms are in most cases characterized as ad hoc terms.

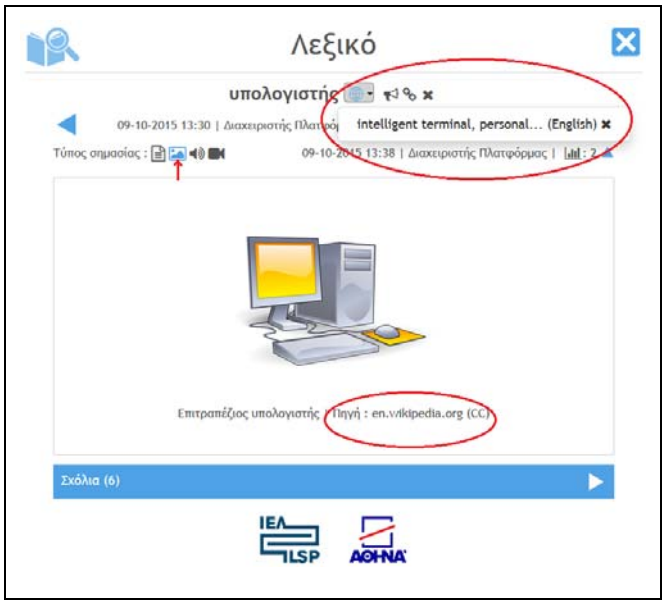

Figure 6: Visualization material associated with the term "computer" is selected.

3.2 Terminology Collection via Crowd Sourcing

Ad hoc terms are those terms translated into a receiver language but not yet undergone the full validation process. Most commonly, ad hoc terms are created within small user groups to serve specific communication needs within these groups. In many cases and for a considerable time phase, there may co-exist parallel versions of ad hoc translations of the same term used in the literature as they derive from different micro-environments, like, for example, different laboratories or university departments in the same national language environment.
This fact alone is a strong deviation from the definition of what a term is, where unique and unambiguous reference are major term properties (Wüster, 1979).
This phase corresponds to step 1 of the validation procedure referred to in section 3.1, and also is the environment from which most usually crowd sourced term collections derive. However, terms collected within this framework still may prove of great value, especially when the need for knowledge communication and understanding is critical as in the case of educational material for deaf student integration in a mainstream university or vocational training environment.
As regards the TEI-A platform, the crowd sourcing option has been a design prerequisite (Tedjamulia et al., 2005), (Wang et al., 2012) aiming to activate volunteers towards the target of collecting as many terms as possible from different sources, in order to facilitate deaf students' integration into the academic environment (Tausczik et al., 2014).
Thus, the platform accommodating the lemmas from the general purpose language sector along with the various terminology collections has been designed to be open to

crow sourcing activities (Doan et al., 2011) for enrichment with new lemma entries.

Following the main *Open eClass* pattern, three basic user roles are supported:

i. Student,
ii. Instructor, and
iii. Administrator.

Individuals from these three categories have equal access rights as regards viewing of educational material, but they also have scalable rights in respect to adding material to the platform. Moreover, the platform also supports special intermediary roles such as "*administrator assistant*", "*user administrator*", "*group leader*" and "*visitor*".

These roles serve, among other functionalities, the goal for lexical material enrichment through crowd sourcing (Sun et al. 2012).

Authorized users may enter new terminology items including the term definition and the whole scale of supporting multimedia material (icons, video, text etc), and they can modify or completely delete entries.

Non authorized users may equally propose new terminology items, being allowed to provide all types of material associated with a term. For security and quality control reasons, however, the items added by non authorized users do not become automatically visible to the whole user community.

Every new addition needs to undergo quality control for content and media (i.e. video quality) before becoming available to the platform user community. Thus, only authorized users, including domain experts from the GSL signers' community, can validate terms suggested by non authorized users and make them visible to the whole user community. It should be mentioned that in this case validation does not correspond to the ISO prescribed procedure, since there is no possibility to go through all term validation steps as foreseen. However, control is performed for content by native GSL signers who are members of the TEI-A community.

The overall terminology enrichment actions open to crowd sourcing initiatives incorporate:

1. New lemma or new sense entry
2. Modification of a lemma or a sense
3. Validation of a proposed lemma sense
4. Communication or hiding of a lemma sense or a lemma description
5. Linking of a lemma with a lemma in a different language (Greek and/or English)

Figure 7: Educational content accessibility support buttons available throughout navigation. Encircled the lexicon activation button.

All information gathered as a result of the above actions is easily accessed via the graphical user interface (GUI) within the lexicon window of the platform. The lexicon window may be activated at any point of the navigation process by clicking on the "book and lens" button (Fig. 7).

4. Conclusion

Here we presented an approach to supporting access of Deaf students to university level educational material via incorporation of GSL elements in a content presentation platform that incorporates elements which serve *Access for All* principles. GSL video presentations of various content parts are supplemented by a lexicon environment incorporating a bilingual general purpose dictionary for the language pair GSL-Greek, and a number of trilingual (GSL-Greek-English) terminology glossaries aimed to support written text understanding. For the enrichment of the terminology glossaries available in GSL, the platform has been implemented following an open design that targets to raising contributions via crowd sourcing.

Since terminology translation validation is a complicated time and human resources consuming enterprise, crowd sourced enrichment of existing resources seems to provide a solution to the severe problem of lack of educational aids in GSL, despite the discussion on ad hoc term status and the more general questioning of crowd sourcing effectiveness (Zhao & Zhu, 2014) after the first period of vast expansion of crowed sourced information collection.

5. Acknowledgements

We acknowledge support of this work by the project "Computational Science and Technologies: Data, Content and Interaction" (MIS 5002437) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

P. Kakoulidis has been a member of ILSP SLT team during the reported platform development period.

6. Bibliographical References

Chamberlain, Ch., Morford, J., Mayberry, R. (1999). *Language Acquisition By Eye*. Psychology Press.

Dimou, A-L., Goulas, T., Efthimiou, E., Fotinea, S-E., Karioris, P., Pissaris, M., Korakakis, D., Vasilaki, K. (2014). Creation of a multipurpose sign language lexical resource: The GSL lexicon database. Proc. of *6th Workshop on the Representation and Processing of Sign Languages,* LREC 2014, Reykjavik, Iceland, pp. 37-42.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. Communications of the ACM, 54(4), 86-96.

Efthimiou, E., Fotinea, S-E., Goulas, T., Kakoulidis, P. (2015). User friendly Interfaces for Sign Retrieval and Sign Synthesis. In M. Antona, M. & C. Stephanidis (Eds.). Proc. of *9th International Conference, UAHCI 2015, Part II*, held as Part of HCI International 2015, Los Angeles, CA, USA, August

2-7, 2015, LNCS vol. 9176, pp. 351-361, Springer, Heidelberg.

Efthimiou, E. "Making Educational Content Accessible for the Deaf: The development of a multi-level platform". In *Proceedings of the 2nd Colloquium on lesser user languages and computer linguistics*, pp: 181-194. Bolzano 13-14 Nov. 2008.

Efthimiou, E. & Fotinea, S-E. (2007). "An Environment for Deaf Accessibility to Educational Content" in *New Trends in Information and Communication Technology & Accessibility, Proc. of ICTA '07*, April 12-14, Hammamet, Tunisia, pp.125-130.

Fotinea, S.-E., E. Efthimiou and A.-L. Dimou (2012). Sign Language computer-aided education: exploiting GSL resources and technologies for web deaf communication *Proceedings of ICCHP-2012 the 13th International Conference on Computers Helping People with Special Needs*, July 2012, Linz, Austria.

Fotinea, S.-E. and Efthimiou, E. (2009). A platform for sign language content presentation. *Proc. of the V International Conference on Multimedia, Information and Communication Technologies in Education* (m-ICTE 2009), April 09, Lisbon, Portugal, pp. 681-685.

Ghari, Z. (2016). The Cognitive, Psychological and Cultural Impact of Communication Barrier on Deaf Adults' Content of Speech in Iran. *Commun Disord Deaf Stud Hearing Aids 4:164*. doi:10.4172/2375-4427.1000164.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition 5/2*, pp:11-30.

ISO/R 704 (1968). *Naming Principles*. International Standard Organization.

ISO/R 1087 (1969). *Vocabulary of Terminology*. Geneva, International Standard Organization.

ISO 704 (1987). *Principles and methods of terminology*. Geneva, International Standard Organization.

ISO 1087 (1990). *Vocabulary of Terminology*. Geneva, International Standard Organization.

ISO 1087/1 (2000). *Terminology work – Vocabulary – Part 1: Theory and application*. Geneva, International Standard Organization.

Katsoyannou, M. & Efthimiou, E. (2004). Terminology theory, methodology and practice. In Katsogyannou, M. & Efthimiou, E. (eds) *Terminology in Greek: Research and Applications*. Kastaniotis, Athens, Greece, pp: 25-60 (in Greek language).

Klima, E., and Bellugi, U. (1979). The signs of language, Harvard University Press, USA, pp.205

Leigh, I. & Andrews, J. (2017). *Deaf People and Society: Psychological, Sociological and Educational Perspectives*. Routledge, New York. (2017, 2nd edition).

Liddell, S. and Johnson, R. (1986). American Sign Language Compound Formation Processes and Phonological Remnants, In *Natural Language and Linguistic Theory,* nol.4, Reideil Publishing Co, pp.445-513.

Phau, R., and Josep, Q., (2010), Nonmanuals: their grammatical and prosodic roles., Sign Languages, In D. Brentari (ed). 381-402. Cambridge: Cambridge University Press.

Sager, J.R. (1994). Terminology: Custodian of knowledge and means of knowledge transfer. *Terminology 1/1*, pp:7-16.

Sandler, W., Lillo – Martin, D. (2006). *Sign Language and Linguistic Universals*, Cambridge University Press, UK, pp.72.

Šumak, B., HeričKo, M., & PušNik, M. (2011). A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types. *Computers in Human Behavior*, 27(6), 2067-2077.

Sun, Y., Fang, Y., & Lim, K. H. (2012). Understanding sustained participation in transactional virtual communities. *Decision Support Systems, 53(1)*, 12-22.

Tausczik, Y. R., Dabbish, L. A., & Kraut, R. E. (2014). Building loyalty to online communities through bond and identity-based attachment to sub-groups. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 146-157). ACM.

Tedjamulia, S. J., Dean, D. L., Olsen, D. R., & Albrecht, C. C. (2005, January). Motivating content contributions to online communities: Toward a more comprehensive theory. In *System Sciences*, 2005. HICSS'05. *Proceedings of the 38th Annual Hawaii International Conference on* (pp. 193b-193b). IEEE.

Wang, H., Chung, J. E., Park, N., McLaughlin, M. L., & Fulk, J. (2012). Understanding online community participation: A technology acceptance perspective. *Communication Research, 39(6)*, 781-801.

Wüster, E. (1979). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexicographie*. (2 Volumes) Wien/New York, Technische Universität Wien.

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 16(3), 417-434.

7. Language Resource References

DIOLKOS (DVD-ROM), 2006. ISBN 960-8191-39-4, http://www.ilsp.gr/diolkos.html