A Digital Moroccan Sign Language STEM Thesaurus

Abdelhadi Soudi[1] & Corinne Vinopol[2]
[1]CLC, ENSMR-Rabat, Morocco
[2]Institute for Disabilities Research and Training, Inc., United States
E-mail: asoudi@gmail.com, corinne@idrt.com

Abstract

This paper presents a gesture-based linguistic approach to assisting Moroccan Sign Language (MSL) users in understanding and appropriately using Science, Technology, Engineering and Mathematics (STEM) terminology by creating the first-ever digital MSL STEM Thesaurus. The thesaurus enables Deaf individuals to describe signs and obtain Standard Arabic word equivalents, concept graphics, and definitions in both MSL and Arabic. This is accomplished not only by providing words comparable to signs that they know, but also by providing other information (e.g., signed definitions) that helps differentiate Arabic word choices. The thesaurus is supported by a Concordancer for better illustration and disambiguation of STEM terms. The thesaurus will likely prove to be an invaluable tool that will enable children and adults who rely on MSL for communication, both deaf and otherwise communication impaired, to better understand and write knowledgeably and clearly on STEM topics, and pass standardized assessments.

Keywords: Moroccan Sign Language, Standard Arabic, STEM Thesaurus

1. Rationale and Background

A population that has been underserved in STEM literacy and under-represented in STEM careers is deaf individuals. There are two primary reasons: First, much of the formal scientific information is not in accessible formats. In fact, most scientific information currently available is audio and text-based and without interpretation into sign language. Second, few educators investigate and use research that points to instructional practices that yield best results for deaf students.

This work addresses the challenging research problem of meeting the educational needs of deaf people who are underserved in education, in general, and in STEM literacy, in particular. In Morocco, there is no secondary education for the Deaf. There is a severe lack of instructional resources: all scientific information is Arabic/French audio and text-based and without interpretation into Moroccan Sign Language (MSL). Second, no materials interpret STEM content into MSL, and likewise, there are no resources that interpret STEM from MSL to Standard Arabic. Many deaf people who rely on sign language for communication do not have good facility with Moroccan Arabic (a spoken language) and Modern Standard Arabic (a written and spoken language, used especially in the media and other professional settings). Since sign languages have no written representation as do oral languages, sign languages can only be represented via video, graphics, and animation. As a result, reading achievement scores of deaf individuals usually fall far short of those found among hearing children of comparable abilities. Studies have shown that the average Deaf adult has the literacy competency of a 10-year-old (Traxler, 2000).

Sign languages can only be described, animated, or videotaped. A few researchers have attempted to develop a notation system to describe individual signs.

Stokoe W.C. (1960, 1965) proposed a notation system for ASL, and Lynn Friedman (1977) provided a phonological analysis of ASL. Contrary to the popular belief, Stokoe realized that signs are not just whole entities, but are composed of smaller atomic units. He developed a transcription system based on sign components which he called “cheremes” and equated with the phonemes of spoken languages. Signs can be described by four cheremes: location, hand shape, motion, and orientation. A number of other writing systems have been developed for representing sign languages in written form. These include HamNoSys (the Hamburg Notational System) Thomas Hanke and Constanze Schmaling (1989) and SignWriting developed by Valerie Sutton (1974). These systems, however, are hardly used or recognized by deaf people or their service providers.

For deaf students, multimedia approaches have been found to enhance factual recall as compared to traditional lecture formats. The combined effects of clear signing, use of media, structured lesson material, and, especially, and interactivity have been found particularly important in terms of performance on post-tests. A study by Dowaliby and Lang (1999) showed that the combined use of signs, graphics, text, and adjunct information also resulted in statistically significant gains as compared to the control group (text only). The results of three different studies with Earth Science, Physical Science, and Chemistry conducted by Donald Steely at the Oregon Center for Applied Science (ORCAS), indicated that interactive multimedia and web-based curriculum materials yielded significantly greater knowledge gains for deaf students as compared to traditional classroom experiences. Lang and Steely (2003) found that well-designed, proven-efficacious science instructional programs for hearing students can be successfully adapted for use with deaf students by interspersing text and American Sign Language explanations with content animation and by providing additional practice on vocabulary and content graphic organizers. Diebold, T. J.

& Waldron, M. B. (1988) concluded that the use of highly pictorial content and simplified English text produced significantly higher pre- to post-test gain scores than formats with less pictorial content.

It is in this context that we have created a STEM thesaurus of MSL and a Concordancing software. This assistive technology will help offer equal access and opportunities to STEM education by providing instructional material.

2. Thesaurus description and functionality

The thesaurus enables Deaf users to:

1. Describe a sign by selecting its cheremes from picture menus;
2. Obtain a graphic and video clip of the sign described by the 8 chosen cheremes;
3. Or, obtain an array of signs that most closely match the user's chereme selections (the chereme version of spell-check);
4. Obtain a list of the Arabic words that can be represented by that sign;
5. Obtain concept graphics to help distinguish the Arabic options;
6. See definitions in Arabic (text) and MSL (video) of the Arabic word options, and
7. Identify word forms and their parts of speech.

The thesaurus creation has been done in two phases. To demonstrate feasibility in Phase I, we picked a small sample of STEM terms to see if we could get the thesaurus to work. It was not a statistically significant sample, just one to check functionality of the software. These signs were selected from the database of software previously developed by our research team (i.e., Sign Generator), which includes 3,000 MSL signs (in both graphic and video format) and 8,500 corresponding Arabic words, symbols, and numbers. To do this, we reordered the database by the sign graphic names. In this way, we could easily determine which signs share more than one Standard Arabic word equivalent. Signs that have more than one Arabic word were given preference for our sample. We then identified the 4 cheremes for each hand for each of these signs (i.e., location, hand shape, palm orientation, and motion for dominant and non-dominant hand) and add their corresponding codes into the database. Each of the 250 chosen signs were given 8 codes, 4 for the cheremes of the dominant hand and 4 for the cheremes of the non-dominant hand. That is, we identified all of the variables for each of the cheremes and then developed a coding system which identified each in the database. For example, if there are 44 hand shapes used in MSL, the "A" hand shape is given the code HS-1, the "B" hand shape HS-2, etc.

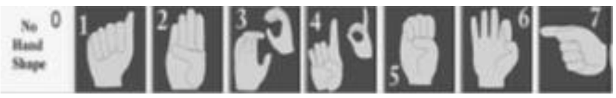

Figure 1: Example of hand shape chereme options and potential codes

Since our existing MSL database did not contain Standard Arabic and MSL definitions for the words that correspond to the 250 selected STEM signs, these were prepared and inputted. Voiceover was also provided for all of the definitions so that it can be appreciated by hearing people who do not have good facility with sign language (e.g., mainstream teachers who do not sign). Audio recording was done separately and then merged with the video before compression. Since the grammars of the two languages (i.e., Standard Arabic and MSL) are divergent, merging them required expertise in both languages (a multidisciplinary deaf and hearing research team).

As is shown in Figure 2, the MSL Thesaurus operates by having MSL users identify the four cheremes for each hand for the STEM sign for which they want to find Arabic equivalents by using drop-down pictorial menus. The program searches the database for the sign that most closely matches the chereme choices. If the cheremes selected do not exactly match how the sign is coded in the database, the program will provide options of signs that are described similarly (the chereme version of spell-check). Accurately described signs and sign options are displayed as graphics and videos. Once users verify their intended sign, the program will display, in addition to the sign graphic and video, the comparable Arabic word(s) and the Arabic word definition(s) in text, MSL definition(s) in video, concept graphic(s), and word forms. This will help deaf students discern, when writing on STEM topics, which Arabic word to use for their sign.

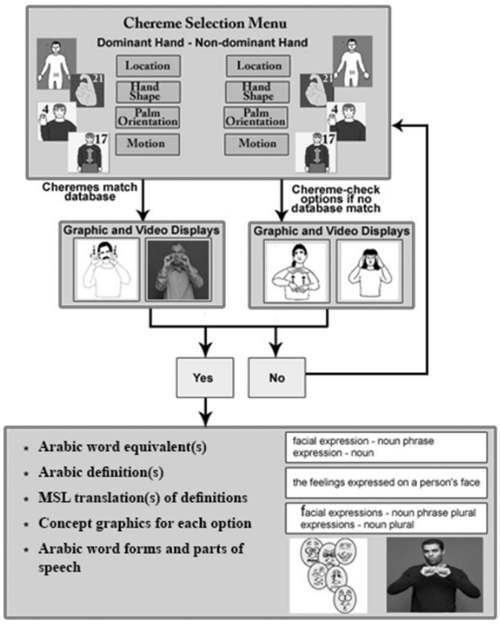

Figure 2: Software navigational functionality
(The sign for "facial expression" is used as an example)

Figure 3 below shows an example resulting from the choice of the 8 cheremes corresponding to the term "friction."

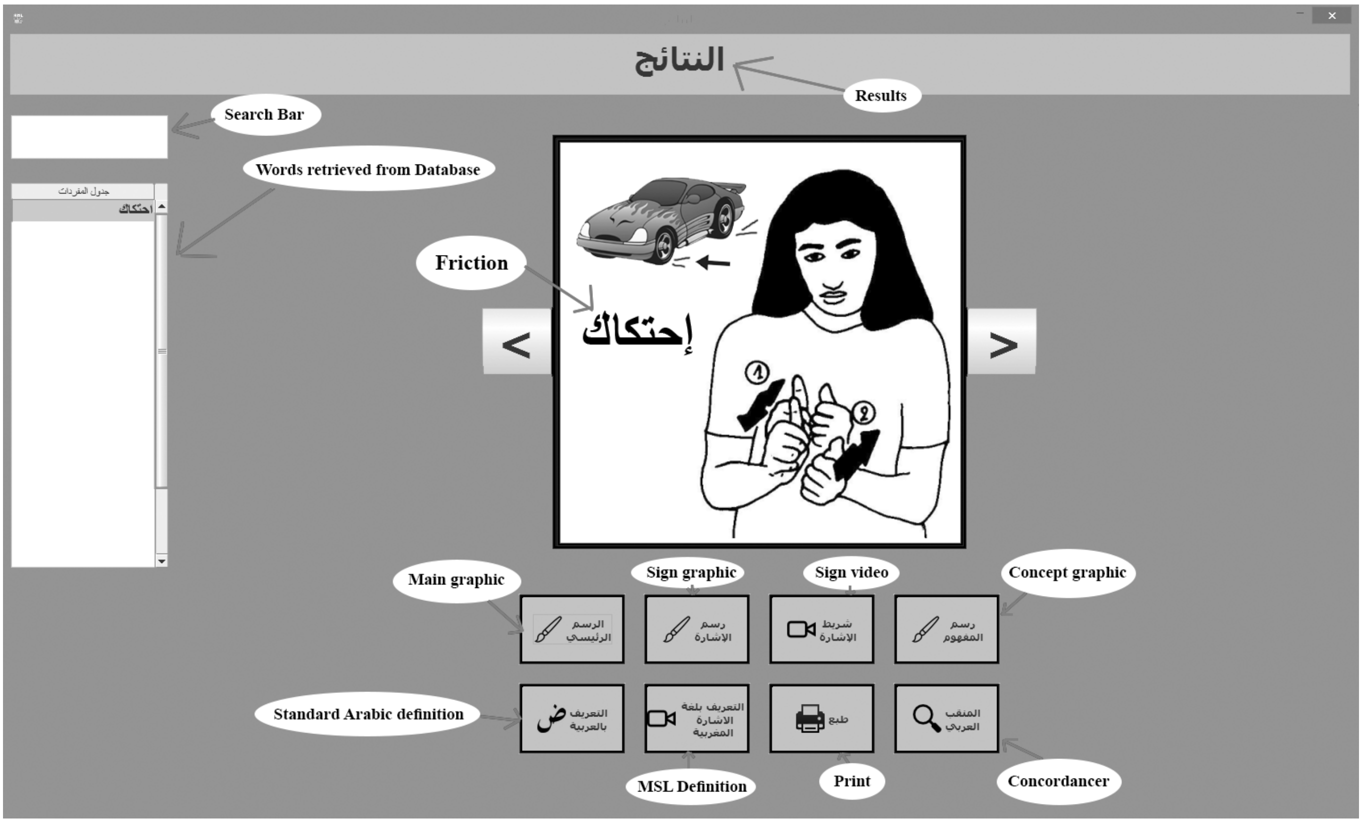

Figure 3: Example of a graphic sign (STEM term "friction") described by the 8 chosen cheremes

Users can choose a variety of output options for the depicted MSL sign (1. Main Graphic Sign which includes the corresponding Standard Arabic term, the concept and the Graphic sign, (2. MSL video clip of the sign, (3. MSL definition, Standard Arabic definition, the concept graphic). As can be seen in Figure 3, the thesaurus is also supported by a Concordancer for a better illustration and disambiguation of STEM terms. This tool provides a list of examples of a particular term or combination of terms, in its/their contexts drawn from a science corpus. By clicking on the Concordancer button, users can are invited to the Concordancer window and are provided a variety of options for searching examples of how the already selected term- "friction" in this case- is used. In order to enable users to search also for other possible inflected and derived forms of a STEM word, we have incorporated Arabic Morphological Analysis in the Concordancer. Arabic morphology/word formation represents a special type of morphological system. It is considered to be a non-concatenative morphology which depends on manipulating root letters in a non-concatenative manner, using different operations such as gemination and infixation. Arabic morphology requires infixation, prefixation and suffixation, giving rise to a large space of morphological variation. Stems are formed by a derivational combination of a root morpheme and a vowel melody; the two are arranged according to canonical patterns. For example, the Arabic stem katab (he wrote) is composed of the morpheme ktb (notion of writing) and the vowel melody morpheme 'a-a'. The two are coordinated according to the pattern CVCVC (C=consonant, V=vowel). This means that Arabic word structure is not built linearly as is the case in concatenative morphological systems. The language has a large degree of ambiguity in word senses, and further ambiguity attributable to a writing system that omits diacritics. (e.g., short vowels, consonant doubling, inflection marks). For example, "ktb" can correspond to kataba «he wrote », kutiba "was written" kutub "books", or 18 other forms). Accordingly, we used a tool that provides all the possible readings/analyses of an inputted word in Arabic. For such a task, we used Buckwater's Arabic Morphological Analyzer (BAMA) (Bucwalter 2002). In BAMA, the data consists primarily of three Arabic-English lexicon files: prefixes (299 entries), suffixes (618 entries), and stems (78, 839 entries)). The tool is based on a concatenative lexicon-driven approach. In (Soudi et al., 2007), we provide a detailed study of Arabic morphological issues.

As is shown in Figure 4, users can choose to see usage examples of the selected term in three ways: 1. examples showing exact match of the term (Figure 4.1), 2. morphological analysis: searching also for different word forms of the selected term. (Figure 4.2), 3. proximity: restricting the search by requiring contexts in which the selected term is adjacent to another specific term chosen by the user (Figure 4.3).

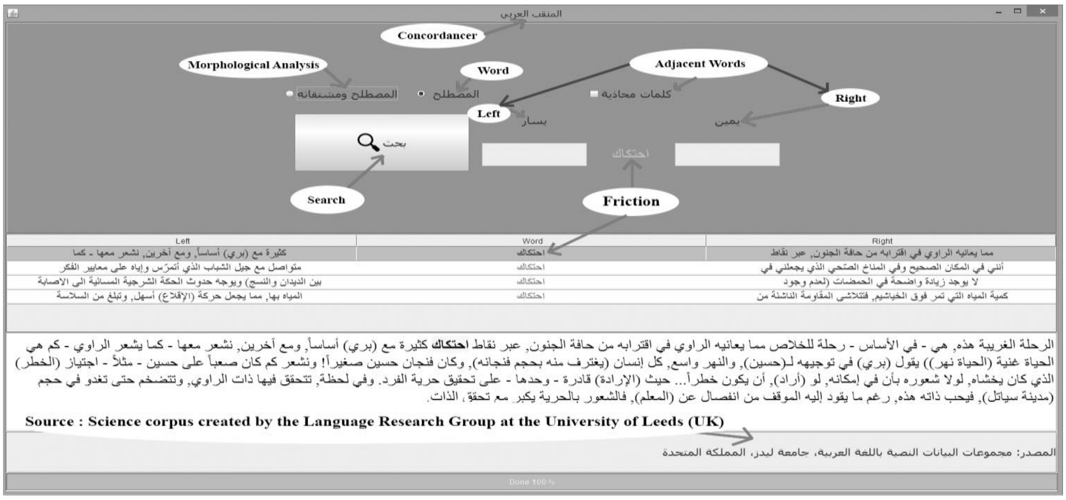

Figure 4.1: Concordancing with exact match search

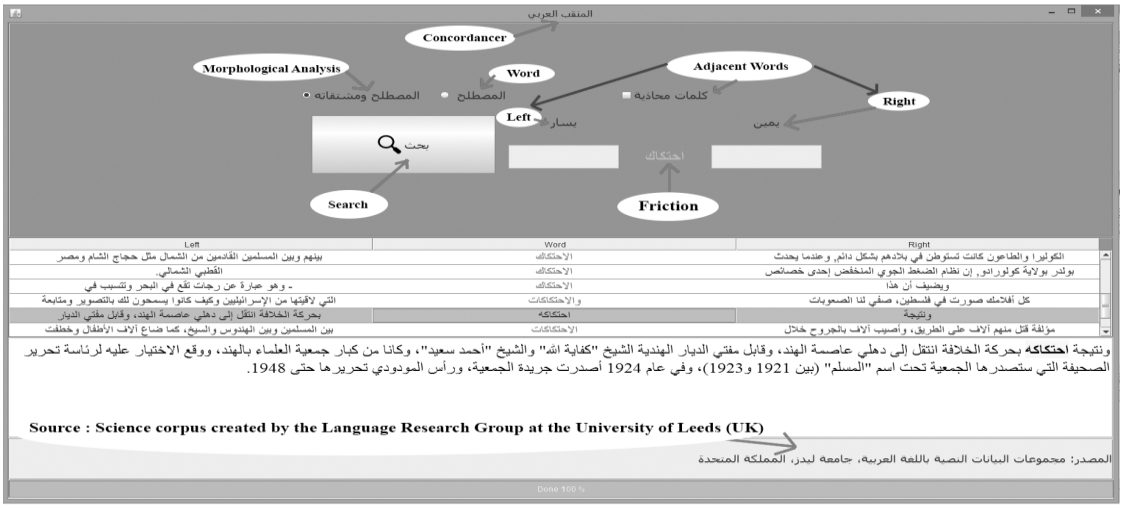

Figure 4.2: Concordancing with morphological analysis

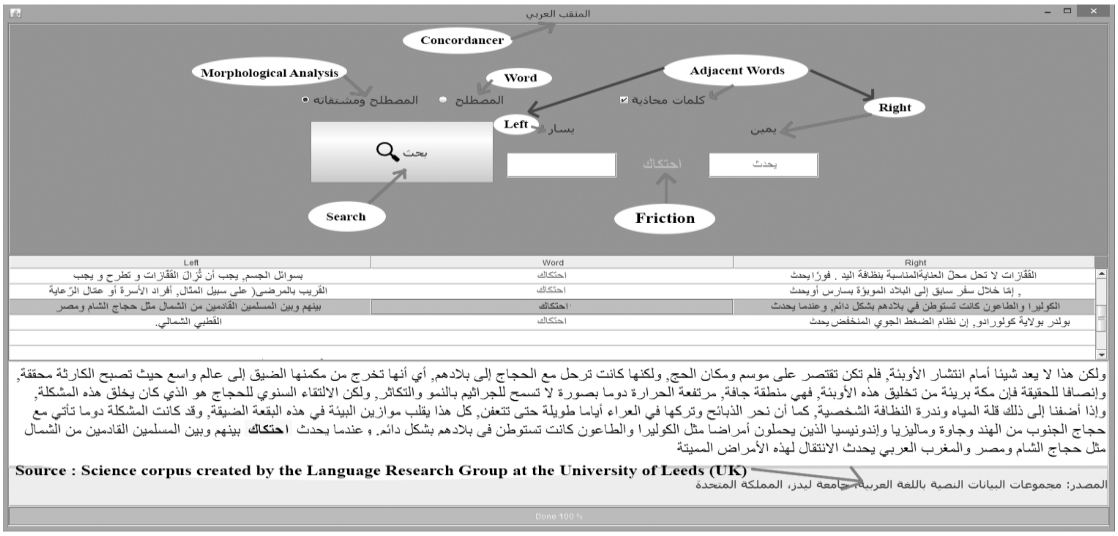

Figure 4.3: Concordancing with proximity

Figure 4: STEM term concordancing options

3. Evaluation of the Thesaurus and its Enrichment

Before proceeding to Phase II (Thesaurus enrichment), some preliminary feedback from typical users was necessary. Twenty deaf and hearing users (including Deaf educators) were invited to test the software in our Sign Language Lab. Ten users were asked to depict 100 STEM terms from four chereme menus to describe the intended sign. Each STEM term had to be depicted within some time frame. The users were able to quickly get exact graphic signs and video clips for 65 STEM terms. An analysis of the chereme variables selected by the users for the other unidentified 35 STEM terms shows that either an exact match is not found, or the user doesn't approve the returned sign and word. This helped us enhance the software so that in cases where the corresponding data is not found or the user does not approve it, the algorithm will return a set of signs that closely match the intended sign. Once the user has approved the sign, the Arabic word equivalent(s), definition(s) (in Standard Arabic and MSL) etc., are fetched from the database and displayed back to the user. In order for us to evaluate the navigational functionality of the software and search efficacy, the other 10 users were given the freedom to describe as many signs (corresponding to STEM terms) as possible by selecting their cheremes from picture menus.

Currently, more in-depth clinical and typical setting usability and efficacy evaluations are being addressed: develop lab observation protocol and usability protocol and arrange evaluation logistics.

The Thesaurus's improvement at the level of both data and navigational features is an ongoing process, and to date, 500 STEM signs are in the database.

4. Conclusion

In this paper, we have described a digital MSL STEM Thesaurus that enables MSL users to describe signs for STEM concepts that they know and use, and find Arabic word equivalents, parts of speech, definitions (in Arabic text and MSL video), and conceptual pictures to help disambiguate meanings. This assistive technology tool will help deaf and hard of hearing students to better understand the nuances of STEM terminology and foster improved written expression to respond to lessons and assessments of STEM content. This is accomplished not only by providing words comparable to signs that they know, but also by providing other information (e.g., signed definitions) that helps differentiate Arabic word choices.

5. Bibliographical References

Al-Sulaiti, L. and Atwell, E. (2006). The design of a corpus of contemporary Arabic. International Journal of Corpus Linguistics, vol. 11, pp. 135-171. [corpus freely available at http://www.comp.leeds.ac.uk /eric/latifa/research.htm].

Battison, R., 1978. Lexical Borrowing in American Sign Language. Linstok Press, Silver Spring, MD (Reprinted 2003, Linstok Press).

Diebold, T. J. and Waldron, M. B. (1988). Designing instructional formats: The effects of verbal and pictorial components on hearing-impaired students' comprehension of science concepts. American Annals of the Deaf, 133, 30-35.

Dowaliby, F., & Lang, H. (1999). Adjunct aids in instructional prose: a multimedia study with deaf college students. Journal of deaf studies and deaf education, 4(4), 270-82. doi:0.1093/deafed/4.4.270.

Friedman, L. A. (Ed.). (1977). Formational properties of American Sign Language. New York: Academic Press.

Lang, H. G., & Steely, D. (2003). Web-based science instruction for deaf students: What research says to the teacher. Instructional Science, 31,277–298.

Ramsey, C., & C. Padden (1998). Natives and newcomers: Gaining access to literacy in a classroom for deaf children. Anthropology and Education Quarterly. 29, 5-24

Singleton, J., Supalla, S., Litchfield, S., & Schley, S. (1998). From sign to word: Considering modality constraints in ASL/English bilingual education. Topics in Language Disorders, 18 (4).

Soudi (2013a). A Robust Technology facilitating interaction between the Hearing Culture and the Deaf Culture. In Proceedings of the session, Interculturalism, Migration and Multiculturalism of the conference "At the Crossroad between Africa, Asia and Europe: Challenges and Perspectives», Venice, Italy, 29 May-01 june 2013.

Soudi, A. Vinopol, C. and Er-rady, A. (2013b). Standard Arabic-to-Moroccan Sign Language Machine Translation: Does translation from a morphologically rich language reflect the interpreters' allocation of high cognitive resources? In Proceedings of The Fourth Conference of the Scandinavian Association for Language and Cognition (SALC IV), University of Eastern Finland, Finland, 12-14 June 2013.

Soudi A., Van den Bosch A. and Neumann G (2007). *Arabic Computational Morphology: knowledge-based and Empirical Methods*. Editors: Soudi A., Van den Bosch A. and Neumann G. In Kluwer/Springer's series on Text, Speech, and Language Technology (series editors Nancy Ide and Jean Veronis).

Stokoe William, Dorothy Casterline, and Carl Croneberg. (1965). A dictionary of American Sign Language. Washington, D.C., Gallaudet College.

Stokoe, William C., Jr. (1960). Sign language structure; an outline of the visual communication systems of the deaf. S.I.L.:O.P.8.

Strong, M., & Prinz, P. (1997). A study of the relationship between American Sign Language and English literacy. Journal of Deaf Studies and Deaf Education, 2(1), 37-46.

Sutton, Valerie. (1974). Sign Writing.

Thomas Hanke and Constanze Schmaling, (1989). HamNoSys, Version 3.0. University of Hamburg, http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html.

Tim Buckwalter. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0. Linguistic Data Consortium, catalog number LDC2002L49 and ISBN 1-58563-257-0. http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49

Traxler, C.B. (2000). The Stanford Achievement Test, 9th edition: National norming and performance standards for deaf and hard of hearing students. Journal of Deaf Studies and Deaf Education, 5, 337–348.

Wilbur, R. and Adamo-Villani, N. (2006-2011). Software for Math Education for the Deaf, NSF award number: 0622900.

World Federation of the Deaf and Swedish National Association of the Deaf's report(2008): Global Education Pre-Planning Project on the Human Rights of Deaf People.