Bilingual resources for Moroccan Sign Language Generation and Standard Arabic Skills Improvement of Deaf Children

Abdelhadi Soudi
Ecole Nationale Supérieure des Mines de Rabat, Morocco
asoudi@enim.ac.ma

Corinne Vinopol
Institute for Disabilities Research and Training
United States
corinne@idrt.com

Kristof Van Laerhoven
University of Siegen, Germany
kvl@eti.uni-siegen.de

Abstract

This paper presents a set of bilingual Standard Arabic (SA)-Moroccan Sign Language (MSL) tools and resources to improve Moroccan Deaf children's SA skills. An MSL Generator based on rule-based machine translation (MT) is described that enables users and educators of Deaf children, in particular, to enter Arabic text and generate its corresponding MSL translation in both graphic and video format. The generated graphics can be printed and imported into an Arabic reading passage. We have also developed MSL Clip and Create software that includes a bilingual database of 3,000 MSL signs and SA words, a Publisher for the incorporation of MSL graphic support into SA reading passages, and six Templates that create customized bilingual crossword puzzles, word searches, Bingo cards, matching games, flashcards, and fingerspelling scrambles. A crowdsourcing platform for MSL data collection is also described. A major social benefit of the development of these resources is in relation to equity and the status of deaf people in Moroccan society. More appropriate resources for the bilingual education of Deaf children (in MSL and SA) will lead to improved quality of educational services.

1 Introduction

Research into brain processing of language has shown that signed and spoken languages occur in the same region of the brain, but there are differences according to language modality of representation (Campbell et al., 2007). This information has implications for designing Natural Language Processing (NLP) systems to facilitate Deaf individuals' access to education. Consideration must be given to language input and output modalities and representations.

Spoken languages rely on audition. Simplistically stated, to express Standard Arabic (SA) in written form, we choose letters that are graphic representations of sounds and put them in that oral production order. When we choose to record or retrieve a word, we can use its spelling and alphabetic sequence.

On the other hand, if a person must rely on vision to develop language, visual principles affect how that language is organized and expressed. Expression tends to be primarily manual and facial, and sign languages (SLs) incorporate many techniques that visually and kinesthetically convey life's experiences. As a result, to express a sign language, it must be depicted through graphics, animation, or video.

Stokoe (1960) coined the terms, chereme and cherology, from the Greek word χείρ for hand. He considered a chereme a basic unit of signed communication, functionally and psychologically equivalent to the phonemes of oral languages. He posited that signs can be described primarily by four cheremes, classified as tab (elements of location from the Latin tabula), dez (the hand shape, from designator), sig (the motion, from signation), and with some researchers, ori (orientation). Facial expression and mouthing are also phonemic in sign language. There have been a few attempts to develop a written form to describe sign language (e.g., SignWriting), but these are hardly used or recognized by Deaf people or their service providers.

2 Deafness and Moroccan Sign Language (MSL)

According to Morocco's High Commission of Planning's 2014 Census (El Ouazzani, 2015) and a survey on disability statistics conducted by the Ministry of Solidarity, Women, Family and Social Development in April 2014 (Lkhoulf, 2017) 3.5% of the population (1,182,681 people) have some degree of hearing loss, 0.2% (56,745 people) have a profound hearing loss or total inability to hear, 1.0% (347,386 people) have "a lot of difficulty"

Proceedings of the 18th Building and Using Comparable Corpora Workshop, pages 1–9
January 20, 2025.

hearing, and 2.3% (778,550 people) have "some" difficulty hearing. There is, therefore, need for both SL use and Deaf education. However, education of Deaf children in Morocco is very dire. Approximately 85% do not attend school. Education beyond sixth grade until recently was unavailable, and very few are gainfully employed. The plight of educating Deaf children is further compounded by other issues (Soudi and Vinopol, 2019).

First, the language of instruction is Arabic/French audio and text-based and without interpretation into MSL unless a volunteer "interpreter" from a Deaf Association is present. Most Deaf children use MSL which is an independent gestural system of communication that does not rely on audition but does, to a great extent, on the logic of the visual experience. It is not an interpretation of SA or spoken vernaculars. It can only be depicted through graphics, video, or animation. Numerous studies have demonstrated that teaching Deaf students is best achieved bilingually (i.e., through both their native signed and spoken languages).

Second, there is a lack of well-trained educators of Deaf children who are familiar with the metacognitive skills essential for effective reading comprehension. Those who can communicate with the children in SL have little training or understanding about how to make educational content meaningful to them.

Third, almost no sign language interpreters exist to help include Deaf children in the regular curriculum with hearing peers. Deaf children who do attend school are kept in segregated classrooms. Therefore, there is little opportunity for Deaf children to get the breadth of educational information that their hearing peers have.

3 Bilingual resources for MSL Generation and SA Skills Vocabulary Building and Improved Reading Comprehension

In view of the challenges outlined above, the development of tools and resources to help Moroccan Deaf children improve their SA skills and access to education is badly needed. In this section, we describe two software programs that we have developed, namely (i) an MSL generator based on rule-based machine translation and (ii) MSL Clip and Create, a set of tools and bilingual resources for custom creation of MSL-supported instructional materials for the improvement of SA skills of Moroccan Deaf children.

3.1 MSL Generator

Several studies have demonstrated that a combination of signed spoken pictures/graphics and comprehension of written text can facilitate Deaf students' spoken language skills and provide support for word recognition (Nielsen et al., 2016; Wilson and Hyde, 1997). Wilson and Hyde (1997) reported that the use of Signed English reading books significantly improves reading comprehension of Deaf students. Similarly, Nielsen et al. (2016) and Stryker et al. (2015) argue that the use of Signed Exact English (SEE) supports the comprehension of reading by Deaf children.

In this subsection, we describe an MSL Generator that uses rule-based machine translation (MT). The system generates sign graphic and video supports for SA text. Teachers can print the graphics and incorporate them into reading passages. Educators of Deaf children can also combine the use of the Generator with the use of the MSL Clip and Create software as described below. That software has sign concepts/images that they can import from the database and add to the reading text.

Research on sign language machine translation (SLMT) is novel compared to research on spoken language MT. Work on SLMT faces some problems, including the lack of parallel corpora, a formal writing system of SLs, and a standard representation format.

Examples of rule-based SLMT systems include the Zardoz system, which translates English text into Japanese Sign Language and ASL (Veale et al., 1998); the Albuquerque Weather System, which translates from English text to ASL in the weather forecast domain (Grieve-Smith, 1999): the TEAM Project (Zhao et al., 2000), which translates English into ASL; and the Greek-to-Greek Sign Language System (Kouremenos et al., 2018).

Data-driven approaches to SLMT include, to name a few, Ebling (2016)'s automatic translation from German to Synthesized Swiss German Sign Language and Bauer et al. (1999)'s statistical-based SL translation system, which translates from recognized video-based continuous SL (German Sign Language) to spoken language (German) in the domain of shopping.

While current SLMT research tends to use data-driven approaches, most (if not all) existing systems either translate in a limited domain or are not actually used by the Deaf Community in real-life situations. This is largely because data-driven ap-

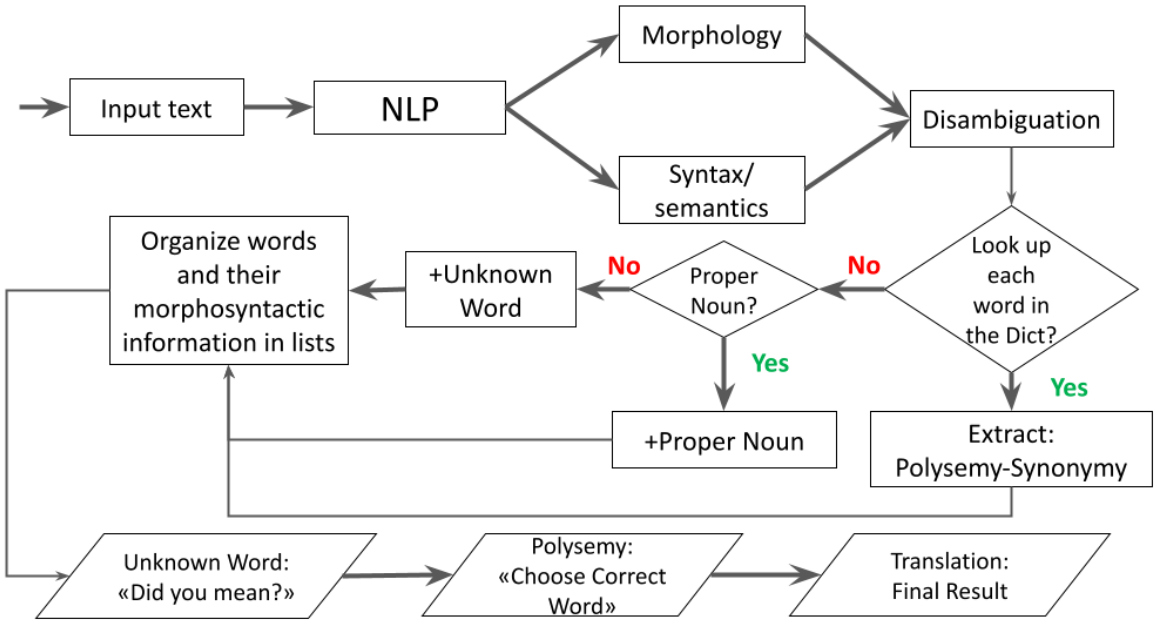

Figure 1: MSL Generator Process Logic

proaches require large-scale parallel corpora that, unfortunately, are not as available for sign languages as they are for spoken languages (Jiang et al., 2023). In addition, most systems do not accommodate SL regional variations.

Using insight gained from previous data-driven SLMT research experiences, we chose to build a rule-based MT (RBMT) system. With this approach, we were able to handle SL regional variations and word sense disambiguation and enrich the current MSL database. Another major reason for choosing a rule-based approach is that it is inclusive of minority or low-resourced languages, such as MSL. Our approach agreed with Hurskainen and Tiedemann (2018, p. 1) who argue that "if we use statistical or neural translation systems, we will exclude 99.8 percent of languages out of development." They further state that "the current hype on neural methods still accelerates the break between the small group of dominant languages and the less-resourced ones. If we want to avoid the break, we do not see any other way out than to put efforts in developing such systems that are affordable for less-resourced languages."

RBMT systems generally handle translation by parsing the source text. The analysis module then results in an intermediate representation that acts as input to the generation module that generates the translation in the target language. Mappings between lexical items stored in bilingual dictionaries and transfer rules are applied to account for the linguistic mismatches between the source language and target language.

Figure 1 shows the MSL Generator logic process. The input to the MSL Generator is Arabic text that is tokenized and then processed using:

(i) Buckwalter Arabic morphological Analyzer (BAMA) (Buckwalter, 2002). BAMA has a lexicon of 40,648 lemmas and three morphological compatibility tables used for controlling affix-stem combinations.

(ii) Stanford parser[1], a statistical parser with pretrained models for English, German, and Arabic.

The MSL Generator then looks up every word in the database. If the word exists in the dictionary, its corresponding sign video and/or graphic (depending on the user's preference) are retrieved. If the word is polysemous, its distinct meanings are also retrieved and displayed on the output screen, each with its corresponding graphic sign. The user then chooses the correct meaning. If the word is not found in the database, the system checks if it can be found in the Named Entity dictionary. If it is not found, the system returns "unknown word" or suggests a possible word(s) based on the context by asking the user "Did you

[1] http://nlp.stanford.edu/software/lex-parser.shtml

Figure 2: Screenshot of the generation of a sentence that includes a polysemous word

Figure 3: Signed Standard Arabic generation after selection of the correct meaning

mean. . . . . ". If all words are found, the user can display the translation of the Arabic text either in the form of graphic sign or video sign.

Figure 2 illustrates the translation of the Arabic sentence كتب الولد الرسالة "The boy wrote the letter." In this sentence, the word "الرسالة" is polysemous. It has three distinct meanings "dissertation," "letter," and "mission." Accordingly, as can be seen in Figure 2, the system displays the three distinct meanings, each with its corresponding graphic sign, and asks the user to choose the intended meaning.

After selection of the correct and intended meaning of the word الرسالة, the system generates the corresponding translation of the input text in MSL, as is shown in Figure 3. Users can also print the sign graphics output.

We are currently developing fragments of MSL grammar through a large MSL corpus that is being created and annotated. We will integrate fragments of MSL grammar into the Generator as much as possible.

We are also incorporating another output mode that will allow MSL Generator to instantly return a pre-made sign language video or avatar sign sequence matching the input text. For this purpose, we use an XML description language (SiGML), which is based on HamNoSys notation (Hanke and Schmaling, 2001), and Lebourque and Geibet's gesture specification language (GessyCA) (Lebourque and Gibet, 1999). A system was developed to convert HamNoSys code of the given word to its SiGML form, to enable the avatar animation.

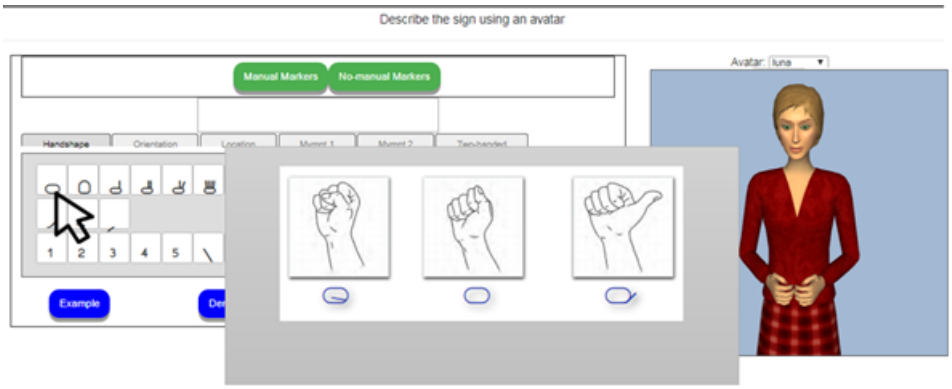

Figure 4: The enhanced version of HamNoSys for the description of signs to animate

To facilitate the animation process, we have created a notation keyboard based on an enhanced version of HamNosys that is legible and user-friendly. HamNosys symbols are enhanced with a pictorial list of the manual and non-manual markers as shown in Figure 4. We are in the process of animating 3,000 SA words in the dictionary used for the MSL Generator.

3.2 MSL Clip and Create Software

MSL Clip and Create is a set of apps, SL media assets, and bilingual resources for the customized creation of MSL-supported instructional material to improve Deaf learners's SA skills. The software includes the following components:

- A bilingual database of 3,000 SA words-MSL signs. Each sign has a corresponding definition in both MSL and SA. Two methods were used to collect data for the database:

  (i) Mobile recording studios: In the first phase, suitable participants from five regions of Morocco (i.e., Souss-Massa, Rabat-Salé-Kénitra, Tanger-Tétouan-Al Hoceima, Marrakech-Safi, and Fes-Meknès) are recruited based on demographic information collected through a detailed questionnaire. The latter includes, inter alia, information on the informants' sex, age, age of MSL acquisition, type of Deafness, frequency of use of MSL, and education. A committee, consisting of a focal team member and local coordinators belonging to a major Deaf association in each region, selects the candidates for the data lexical elicitation task. Following Schembri et al. (2012), Milroy (1980), and Bayley et al. (2001), we involve local Deaf signers from regional associations to both recruit local informants and lead the data collection process. This method ensures that the data collection tasks are Deaf-friendly and adapt to the sociocultural experiences of the participants. The importance of involving Deaf assistants in the process of data collection has been documented in the literature (Harris et al., 2009; Ladd, 2003; Singleton et al., 2012). These authors also recommend that the required informed consent be translated into the participants' native language. This recommendation is particularly necessary in Morocco, where the Deaf community has very limited spoken language proficiency. Using this method, we were able to initially collect 2,200 signs.

  (ii) In order to accommodate regional sign variations across Morocco and to meet the needs of educators of Deaf children, we developed a crowd-sourcing platform, Madrasati-Signs «My school signs» [2] that includes a database of 3,000 words categorized into 21 domains (i.e, family, colors, home, food, clothes, time, sports and hobbies, body, feelings, geography, transportation, education, nature and weather, animals, health, math, business and careers, media and arts, technology, alphabet, life sciences and physics). The choice of words is based on an enhanced version of MacArthur-Bates Communicative Development Inventory Words and Moroccan (STEM)

[2] https://madrasati-signs.org/

textbooks (Fenson et al., 2007). Most words in the database have a corresponding concept. The participants were asked to record signs for words (and concepts) in a specific domain using the Madrasati-Signs platform. The platform is designed in a similar way as the crowdsourcing platform AfricaSign (Soudi et al., 2019). Madrasati-Signs can accommodate the recording of multiple signs (if any) for each word[3]. Where regional variations exist for a single sign, they are identified as (sign 1), (sign 2), and so forth. Users have two input modes:

a. Add their signs by videotaping them using Laptop/Phone cam. After their consent, the users' phone/laptop cams will be automatically activated, and they will be asked to provide a sign for a particular word. They will then have the possibility of viewing their sign, and either validate it or videotape it again. Figure 6 shows a Deaf signer from Rabat recording signs in the Home domain.

b. Uploading a video sign if a user already has it.

For quality assurance purposes, sign contributions to the platform were restricted to trusted signers selected by regional Deaf associations. Madrasati-signs users logged in and described themselves demographically by region and Deafness affiliation (e.g., Deaf themselves, have Deaf parents-CODA).

- MSL Clip and Create also includes a publishing tool for the creation of customized and printable materials using the graphics in the database. Users can also import other graphics and photos from their device. Educators of Deaf children, for example, can import a SA reading passage from a Moroccan textbook and support it with graphic signs and concepts from the database, as is shown in Figure 5. The user can import as many sign graphics as they desire in real time. This tool is particularly useful in education environments in which textbooks are designed for the hearing and are not adapted to the needs of Deaf learners, as in the case of Morocco.

Figure 5: A screenshot of the Publisher illustrating an imported reading passage on Nature with graphic sign supports from the database

- In addition to the database, and publisher, the software also has six templates that instantly create bilingual SA-MSL customized crossword puzzles, word searches, bingo cards, matching games, flashcards, and fingerspelling scrambles using any graphic signs of the database. Figure 8 shows a screenshot of a customized bingo card and fingerspelling scramble puzzle.

- The software also includes a story-builder that currently hosts three Hispanic folktales translated into SA and three stories from a Moroccan national SA textbook. Users can view the story in MSL and/or read the Arabic text. The latter can be automatically diacritized with a simple click of a button. This functionality is necessary for early grade children who still do not know the grammar of SA and, therefore, cannot read it without diacritics. Figure 7 shows a screenshot of one of the stories in the MSL Clip and Create software.

These tools and resources can be used for

[3]For the list of words and their corresponding concepts that were included in the lexical elicitation task, see www.madrasati-signs.org

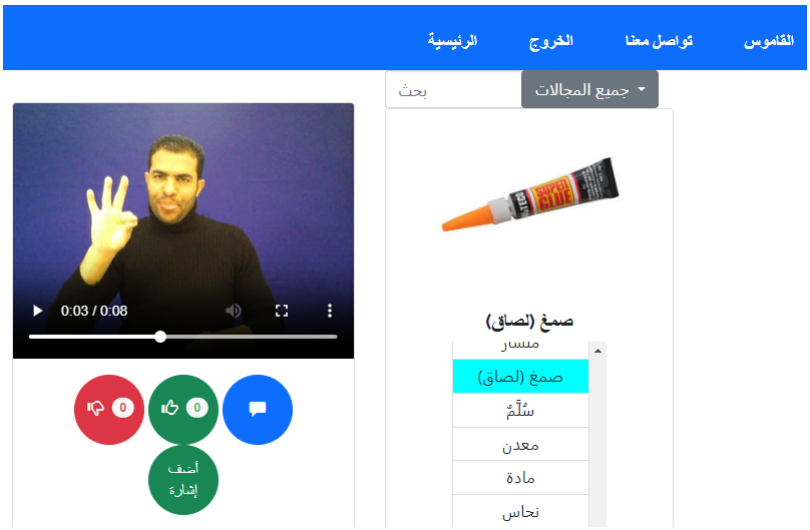

Figure 6: A screenshot of a Deaf signer from Rabat recording signs in the Home domain using Madrasati-Signs platform

Figure 7: A screenshot of one of the Software's stories

reading-related activities, such as SA vocabulary building and word recognition in each of the 21 dictionary domains.

4 Broader Impact

In addition to academic benefits, a major social benefit of the development of these MSL-SA resources is in relation to equity and the status of Deaf people in Moroccan society. More appropriate resources for the bilingual education of Deaf children (in MSL and Arabic) will lead to improved quality of educational and interpreting services for Deaf people and provide more opportunities for self-development and employment. Deaf people who can become more highly qualified and trained will be in a better position to contribute to society in different ways, and will be able to achieve greater recognition, access, and equity in the wider community. Furthermore, the greater understanding of MSL and improved resources for MSL teaching, learning and research can provide an evidence-base for policy-makers in supporting appropriate education, training and services for Deaf children and adults. In this context, it is worth noting that the Moroccan Ministry of Education has endorsed our tools and resources and helped with their free distribution to Deaf associations across Morocco. The Ministry is currently investigating establishment of a teacher of the Deaf training program. These efforts will help close the gap in education, employment, and health between Deaf people throughout their lifespan and their hearing peers.

5 Limitations

As is the case of most other languages, one of the major limitations of the MSL Generator is the lack of a comprehensive grammar of MSL. SLs are natural languages, and they have also developed linguistic systems with a grammar and a vocabulary (Johnston and Schembri, 2007). However, there is no other SL that has a reference grammar "that meets

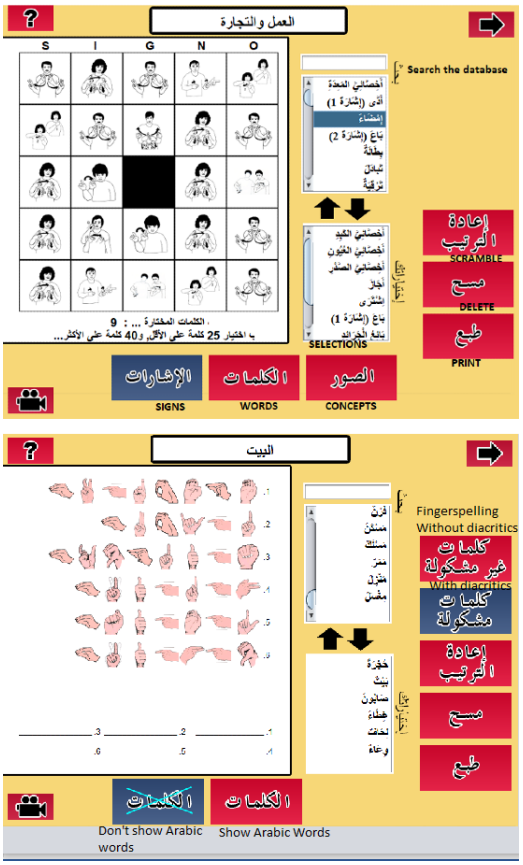

Figure 8: Screenshots of bingo and fingerspelling templates

the common standards set by spoken language reference grammars" (Palfreyman et al., 2015). We are currently addressing this limitation though the creation of a large scale MSL corpus which is an important resource to understand the grammar of MSL. Data is being collected from a total of 240 Deaf signers from the twelve regions of Morocco. The dataset, totaling 120 hours of video, is being/will be translated, of which 10 hours are being annotated, and tagged using EUDICO Linguistic Annotator (ELAN). Fragments of MSL grammar will be incorporated into MSL Generator. The corpus creation will also enrich the SA-MSL database and help understand the sociolinguistic situation of MSL by investigating factors, such as the multilingual linguistic environment, gender, regional variation, family and education.

6 Conclusion

This paper has presented a set of SA and MSL tools and resources calculated to improve Moroccan Deaf children's SA skills. We have described two self-developed resources that are intended to help Deaf learners improve their SA skills. MSL Generator enables educators of Deaf children to enter Arabic text and generate its corresponding translation in both MSL graphic and video formats. The generated graphics can be printed and imported into an Arabic reading passage. We have also described MSL Clip and Create software, which includes a database of 3,000 SA words and MSL signs, a Publisher for the incorporation of MSL support into Arabic reading passages, and six Templates that create customized crossword puzzles, word searches, Bingo cards, matching games, flashcards, and fingerspelling scrambles.

Helping Deaf children improve their SA skills is challenging and requires a strong long-term commitment, particularly in light of the lack of resources available in their native sign language.

References

Britta Bauer, Sonja Nießen, and Hermann Hienz. 1999. Towards an automatic sign language translation system. In *Proceedings of the international workshop on physicality and tangibility in interaction: towards new paradigms for interaction beyond the desktop, Siena, Italy*.

Robert Bayley, Clayton Valli, and Ceil Lucas. 2001. *Sociolinguistic variation in American sign language*. Gallaudet University Press.

Tim Buckwalter. 2002. *Buckwalter Arabic Morphological Analyzer Version 1.0*. Linguistic Data Consortium, University of Pennsylvania. LDC Catalog No.: LDC2002L49.

Ruth Campbell, Mairéad MacSweeney, and Dafydd Waters. 2007. Sign language and the brain: a review. *Journal of deaf studies and deaf education*, 13(1):3–20.

Sarah Ebling. 2016. *Automatic Translation from German to Synthesized Swiss German Sign Language*. Ph.D. thesis, University of Zurich.

Z. El Ouazzani. 2015. Moroccan experience on disability statistics. Washington Group Meeting, Copenhagen, Denmark. Retrieved from https://www.cdc.gov/nchs/data/washington_group/meeting15/wg15_session_8_4_touahami.pdf.

Larry Fenson et al. 2007. Macarthur-bates communicative development inventories.

Angus B Grieve-Smith. 1999. English to american sign language machine translation of weather reports. In *Proceedings of the Second High Desert Student Conference in Linguistics (HDSL2), Albuquerque, NM*, pages 23–30.

T Hanke and C Schmaling. 2001. A hamnosys-based phonetic transcription system as a basis for sign language generation. In *Gesture Workshop 2001*.

Raychelle Harris, Heidi M Holmes, and Donna M Mertens. 2009. Research ethics in sign language communities. *Sign Language Studies*, 9(2):104–131.

Arvi Hurskainen and Jörg Tiedemann. 2018. Rule-based machine translation from english to finnish. In *Proceedings of the Second Conference on Machine Translation (WMT2017)*. The Association for Computational Linguistics.

Zifan Jiang, Amit Moryossef, Mathias Müller, and Sarah Ebling. 2023. Machine translation between spoken languages and signed languages represented in SignWriting. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1706–1724, Dubrovnik, Croatia. Association for Computational Linguistics.

Trevor Johnston and Adam Schembri. 2007. *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.

Dimitrios Kouremenos, Klimis Ntalianis, and Stefanos Kollias. 2018. A novel rule based machine translation scheme from greek to greek sign language: Production of different types of large corpora and language models evaluation. *Computer Speech & Language*, 51:110–135.

Paddy Ladd. 2003. *Understanding deaf culture: In search of deafhood*. Multilingual Matters.

Thierry Lebourque and Sylvie Gibet. 1999. High-level specification and control of communication gestures: The gessyca system. In *Computer Animation, 1999. Proceedings*, pages 24–35.

N. Lkhoulf. 2017. Disability statistics from the 2014 moroccan census. In *Regional Meeting on Disability Measurement and Statistics in support of the 2030 Agenda for Sustainable Development and the 2020 World Population and Housing Census Programme*, Muscat, Oman.

Lesley Milroy. 1980. Language and social networks. *(No Title)*.

Diane Corcoran Nielsen, Barbara Luetke, Meigan McLean, and Deborah Stryker. 2016. The english-language and reading achievement of a cohort of deaf students speaking and signing standard english: A preliminary study. *American Annals of the Deaf*, 161(3):342–368.

Nick Palfreyman, Keiko Sagara, and Ulrike Zeshan. 2015. Methods in carrying out language typological research. In Eleni Orfanidou, Bencie Woll, and Gary Morgan, editors, *Research Methods in Sign Language Studies: A Practical Guide*, pages 173–192. Wiley-Blackwell, Chichester, UK.

Adam Schembri, Jordan Fenlon, Ramas Rentelis, Steve Reynolds, and Kearsy Cormier. 2012. Towards a british sign language corpus: A short report. *International Journal of Corpus Linguistics*, 17(1):3–15.

Jenny L Singleton, Gabrielle Jones, and Shilpa Hanumantha. 2012. Toward ethical research practice with deaf participants. *Journal of Empirical Research on Human Research Ethics*, 9(3):59–66.

Abdelhadi Soudi, Kristof Van Laerhoven, and Elmostafa Bou-Souf. 2019. Africasign – a crowd-sourcing platform for the documentation of stem vocabulary in african sign languages. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, page 658–660, New York, NY, USA. Association for Computing Machinery.

Abdelhadi Soudi and Corinne Vinopol. 2019. Educational challenges for deaf and hard-of-hearing children in morocco. *Deaf education beyond the western world: Context, challenges, and prospects*, pages 307–322.

William C. Stokoe. 1960. Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of Deaf Studies and Deaf Education*, 10(1):3–37.

Deborah Stryker, Diane Nielsen, and Barbara Luetke. 2015. Signing exact english: Providing a complete model of english for literacy growth.

Tony Veale, Alan Conway, and Bróna Collins. 1998. The challenges of cross-modal translation: English-to-sign-language translation in the zardoz system. *Machine Translation*, 13:81–106.

Tamara Wilson and Merv Hyde. 1997. The use of signed english pictures to facilitate reading comprehension by deaf students. *American Annals of the Deaf*, pages 333–341.

Liwei Zhao, Monica Costa, and NL Badler. 2000. Interpreting movement manner. In *Proceedings Computer Animation 2000*, pages 98–103. IEEE.