Multilingual Synthesis of Depictions through Structured Descriptions of Sign: An Initial Case Study

John McDonald[1,2], Eleni Efthimiou[2], Stavroula-Evita Fotinea[2], Rosalee Wolfe[1,2]
[1] School of Computing, DePaul University, Chicago, IL, USA
[2] Institute for Language and Speech Processing, ATHENA Research Center, Athens, Greece
{jmcdonald, rwolfe}@cs.depaul.edu, {eleni_e, evita, rosalee.wolfe}@athenarc.gr

Abstract

Sign language synthesis systems must contend with an enormous variety of possible target languages across the world, and in many locations, such as Europe, the number of sign languages that can be found in a relatively limited geographical area can be surprising. For such a synthesis system to be widely useful, it must not be limited to only one target language. This presents challenges both for the linguistic models and the animation systems that drive these displays. This paper presents a case study for animating discourse in three target languages, French, Greek and German, generated directly from the same base linguistic description. The case study exploits non-lexical constructs in sign, which are more common among sign languages, while providing a first step for synthesizing those aspects that are different. Further, it suggests a possible path forward to exploring whether linguistic structures in one sign language can be exploited in other sign languages, which might be particularly helpful in under-resourced languages.

Keywords: Sign Language Synthesis, Avatar, AZee, Geometric Constructions, Multilingual, Translation

1. Introduction

Signing avatars have held an, as of yet, unrealized promise both as an assistive technology for bridging Deaf-hearing communication and as an educational tool for Deaf and hearing sign language learners. Even while advances in both computer animation and machine learning are bringing us closer to realizing some of these long-held goals avatars are often eyed with suspicion by the Deaf community. This is due to their failure to legibly portray both the full linguistic structure of signing and the subtleties of human motions (Kipp et al., 2011). Another contributing factor are overly confident claims often made by companies and researchers concerning the capabilities of their avatars (Wolfe et al., 2022) (Deutscher Gehörlosen-Bund et. al., 2024).

One significant challenge for wide-scale applicability of signing avatars is the great diversity of sign languages across the world. Ironically, it is a common misconception among the hearing population that sign languages must be "universal", partially because of the perceived prevalence of iconicity in signing (Hohenberger, 2007). People in the U.S., for example are often surprised to hear that, not only is British Sign Language (BSL) a completely separate sign language from American Sign Language (ASL), but that ASL shares more in common with French Sign Language (LSF) from which it was derived (Fischer, 2015).

More surprising yet is when they hear that Switzerland has three recognized sign languages, Swiss German (DSGS), Swiss French (LSF-SR), and Swiss Italian (LIS-SI), used in different regions and are different both from each other and from the sign languages of Germany (DGS), France (LSF) and Italy (LIS) (Eberhard et al., 2022). A similar situation can be found across South America where Venezuelan, Honduran and Argentine sign languages are all distinct despite the fact that Spanish is the common spoken language across the region (Akorbi, 2023).

An avatar that signs in only one language will be limited to serving the population of a single region, and several projects have worked towards creating a signing avatar that can communicate in several languages (Efthimiou et al., 2010). However, each of these have been limited in the linguistic features that they can encompass and the naturalness of human motion that they can achieve, both of which decrease the legibility of the resulting synthesized sign. More recently, the EASIER project has taken up this challenge and the present case study has arisen from this work (EASIER-Project, 2024).

Throughout this discussion it will be important to remember that a signing avatar is not the same as a spoken-to-sign translation system. The same is true for a linguistic description system. While it is true that considerations of Deaf-hearing communication cause researchers to focus more often on translations of spoken and written language to sign, most signing happens among native signers in the Deaf community and has no relationship to spoken or written language, nor is there a broadly accepted written form for sign language. It is imperative then to see both the linguistic description and the signing avatar for what they actually are, respectively, a description of the signed discourse, and a display system to communicate that signing visually. Whether the signed discourse arises

Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages, pages 244–253

from a spoken/written translation or from naturally occurring sign is irrelevant.

This paper will present promising first steps towards animating sign in multiple languages directly from a rich structured description of the desired signing. The descriptions are not tied to translations from spoken languages and can also encompass natural signing that arises between native signers. The hierarchical structure afforded by the linguistic descriptions provides important cues to the animation system that informs nonmanual and prosodic signals which include the relative timing of both manual and non-manual motions (Sandler, 2010). These significantly improve the legibility of the resulting synthesized sign. Furthermore, this effort focuses on forms signing that have traditionally been a challenge for sign synthesis systems, namely classifier constructions and geometric depictions. In fact, it is the very geometric nature of these constructs that makes them more understandable among different sign languages and provides a foundation upon which a multilingual system may be achieved.

Finally, these first steps will point to important ways that linguistic descriptions of sign and avatar animations can interact, allowing each side to learn from the other. Indeed a synthesis system built on hierarchical descriptions of sign and legible animation may provide a powerful tool for studying and testing linguistic theory.

2. Lessons from prior multilingual efforts

A brief review of past efforts towards the multilingual display of sign language can highlight many of the challenges faced in such an endeavor, and can also point to strengths in each approach that can be leveraged in the new approach explored here. In the past, the main efforts for sign language representation and display can be divided into three main categories:

1. *Phonetic systems* such as HamNoSys attempt to encode the motions of signing via the fundamental parameters of human posture and motion, such as handshape, palm orientation, movement (Hanke, 2004). For example, large libraries of HamNoSys/SigML annotations were used in the Dictasign project to allow signing in several European sign languages (Dictasign, 2012). Dictionary signs and other gestural units were described to the avatar phonetically in terms of their parametric linguistic labels to enable easy annotation. The supporting avatar was quite flexible in the range of vocabulary it could express, due to the existence of large corpora of annotated signs in a few languages. However it was limited in the quality of animation output due to the coarseness of the linguistic description, and the lack of prosodic cues included in the animation (Caridakis et al., 2011) (Kipp et al., 2011). Structure beyond the phonetic is necessary for portraying the prosodic structure of the language, which is essential for legibility. Nevertheless, phonetic linguistic notations can provide large repositories of data for an animation system.

2. *Gloss-based systems* rely on a series of glosses, i.e. written words that provide the closest approximation to the meaning of a sign, which dictate the content of the desired signing. A dictionary-based lookup of these concepts from the target sign language is then used. This lookup can be in the form of preanimated (Wolfe et al., 2011) or prerecorded sequences (Gibet and Marteau, 2023). While very flexible, these efforts suffer from small vocabulary sizes, due to the cost of either animating sign or recording humans with motion capture. In addition, while the signing of individual dictionary entries can be of very high quality in either approach, the process of stitching sequences of these recordings can be stale and awkward if the system has no knowledge about the larger grammatical structures that link them together. This includes both nonmanual and timing considerations. Because of this, some more recent gloss based systems have explored adding prosodic and nonmanual instructions to gloss streams (Adamo-Villani and Wilbur, 2015) (Hanke et al., 2023); developments which have greatly enhanced the quality of the resulting animation. One of the great lessons from this approach is that the more structure that the representation provides, particularly for prosodic and nonmanual communication, the more legible the synthesized sign will be. In addition, using a library of phonetic description such as HamNoSys to describe each gloss to the avatar could help alleviate the problem of small dictionaries, but at present efforts to animate directly from such sparse linguistic data remains problematic due to the robotic nature of the resulting motion.

3. *Deep-learning systems* which exploit large libraries of annotated video and/or motion capture recordings of sign, and attempt to produce video or skeletal motion directly from the desired spoken text (Saunders et al., 2021). These efforts have explored multilingual display in British Sign Language (BSL), the Sign Language of the Netherlands (NGT) and DGS. The major current challenge for these tech-

niques is the size of available corpora. Among the largest annotated corpus between signed and spoken languages is the DGS Corpus (Hanke et al., 2020), which contains in excess of 63,000 pairs. This may seem large, but pales in comparison to the roughly 15 billion pairs that are exploited for modern translation systems between spoken languages. These efforts also suffer from a major issue when it comes to linguistic study. Since the neural networks that drive these systems are largely black-boxes that produce an animation with little indication of how the system is producing the result, it can be difficult to derive meaningful linguistic data on the structure of the resulting signed discourse. Nevertheless, there is no denying the power of deep learning techniques, and as both the corpora and the techniques that exploit them advance, they will no doubt provide increasingly important for informing animation tasks (Choudhury, 2022).

This last issue of corpus size can be particularly problematic for so-called under-resourced sign languages. Compared to spoken languages, most NLP and Deep Learning efforts would consider all sign languages under resourced (Börstell, 2023), however, even among sign languages there is a great disparity between the resources amassed for languages like ASL, DGS, and BSL, and those that have been gathered for languages with smaller communities such as Greek Sign Language (GSL), for which corpus sizes and native populations are significantly smaller. Any effort that aims to deal with multilingual display must have a method for handling such disparities.

3. Challenges and opportunities

The misconception that sign language should be universal does arise out of two interesting aspects of both sign languages and their native users. First, there is evidence that signers, as opposed to users of spoken language, are often more adept at interlingual communication. One factor may be due to the continual practice signers get when attempting to communicate with both hearing people and Deaf people from other cultures (Sacks, 2022).

Another important factor is the very nature of the languages that they are fluent in. There tend to be more aspects of signing that are found to be shared between sign languages than is the case for spoken languages. One of these is the form in which an utterance can mimic the shape, motion or sound that is being conveyed (Perlman et al., 2018). [1]. The visual structure of sign includes geometric constructs such as classifier predicates, size and shape specifiers and and depicting signs that use the body in geometric ways (Zwitserlood, 2012).

In fact, it is partly due to the prevalence of certain types of geometric constructs, which are similar across sign languages, that has led to the greater success of International Sign compared to similar efforts in spoken language (Mesch, 2010). In particular, International Sign often uses classifiers and depictions as a more interlingually understandable way to communicate objects and actions than fixed signs (McKee and Napier, 2002). The present effort will seek to exploit these aspects in an effort to build a first step towards a multilingual display.

3.1. Classifiers and depictions

Geometric constructions, including size and shape specifiers and classifier predicates that depict the placement and movements of objects are observed in most sign languages, and, while the specific handshapes differ significantly between sign languages, the motions of the body that depict the placement and movement are largely similar (Pfau et al., 2012). For example, when placing a small round object like a plate, many languages use the hands to mimic the shape of the object, and then use a downward motion to place that object figuratively in space relative to other objects that may be depicted. Unless the object is hanging on a wall or on the ceiling, this motion will naturally be downward. More generally, the placement of the object will be expressed naturally by a movement toward the surface that the object is resting on.

Extensive examples of geometric constructs like these may be found in the Mocap1 corpus in LSF from the LISN (formerly LIMSI) laboratory (LIMSI et al., 2022) (Benchiheub et al., 2016). In this corpus, Deaf participants were provided with pictorial stimuli that they were then free to describe in any way they wish. For example, one stimulus was the picture of the neatly-decorated dining room shown in figure 1. Descriptions of the room varied significantly between participants with some describing the room very sparsely and others in great detail.

One key characteristic of classifier constructions in sign language is that they are among the least "lexical" parts of signing that occur in native discourse in the sense that signers will often use very few dictionary signs when describing either the action or structure of a scene. In a 30-second section of one Mocap1 participant's description, the signer described the table setting in great detail. The entire sequence, however involves only seven dictionary signs, which in English would be glossed RUG, TABLE, CHAIR, PLATE, GLASS, KNIFE and

[1] In spoken language onomatopoeia is an example of iconicity, and some sources (Perlman et al., 2018) (Handspeak, 2) have applied versions of the term onomatopoeia to signing as well in place of iconicity

Figure 1: Mocap1 stimulus for the description of a dining room

FORK. The rest of the signing is dedicated to the geometric placement of the rug on the floor, the table on top of the rug, two pairs of chairs facing each other at the table, four plates arranged symmetrically on top of the table with the four glasses and the four pairs of knives and forks arranged around the plates. Figure 2 contains examples of the signer placing a plate, two glasses and a knife and fork pair.

It is precisely due to the fact that classifier predicates show a geometric placement or movement of an object that, once the object's type is established, the action mimics the natural ways in which the object settles onto a surface or moves in space. This makes it far more likely that these motions will be similar among sign languages. Note, however, that it is not claimed here that they are precisely the same in all sign languages, but that their similarity gives a starting point to work from for multilingual display. The system described below will have the flexibility to accommodate such differences. For example, the classifier for a moving vehicle like a car is signified in ASL by an "Three" handshape oriented on its side with the ulnar side of the palm against the surface, whereas in LSF, it is indicated by a flat hand with the palm flat on the surface. In both instances, the extended fingers indicate the direction that the car is facing, though this may not always be the case for other language pairs[2].

3.2. Sign language description

Every synthesis system must have, at its core, some method of describing the desired signing, and we must consider the lessons from prior efforts discussed in section 2, when choosing the description system. Of primary consideration for our case study here is the legibility of the resulting synthesis, and as has been seen in prior work, non-manual signals that give purposeful motion to the spine, head and face are key to the quality of the resulting signing. Further, the description system must be capable of informing varied timing and pacing for movements that are key to breaking up the robotic monotony that have plagued past efforts to synthesize sign from phonetic descriptions. In this respect, the AZee description system (Filhol et al., 2014) has proven to be a powerful tool for describing, not only the basic gestural units of signing, but also the connecting structure that provides necessary nonmanual and timing information (Filhol et al., 2017).

Another key aspect of AZee from a synthesis perspective is that it has proven extremely capable in its ability to describe classifier constructions and depictions for avatar synthesis, where this table description was animated directly from the AZee description in LSF (McDonald and Filhol, 2021). The only elements that were supplied by an artist were the seven animations of the citation forms of RUG, TABLE, etc, and single example poses for each of the classifiers. We will not review the AZee description in its entirety, but will recall the main AZee rules that are used in the description, and select examples of how those rules are used. The names and parameters of the rules have been updated in accordance with the latest published AZee notation (Filhol et al., 2024).

- *in-context*(*context*, *process*), formerly *context*, this rule provides the main glue that knits sections of signing together. It causes a hold to happen at the end of the *context* along with a blink. It indicates that the signing described by *process* is to be understood in the context of the signing *context* that comes before it. For example,

  in-context(*table placed on rug*, *items on table*) (1)

  indicates that the table placed in the scene is where the list of items is placed. This is often one of the top-level rules that builds the hierarchy of signed discourse, and both *context* and *process* are often large descriptions of signing themselves.

- *instance-of*(*type*, *element*), formerly *category*, this rule indicates that the signing described by *element* is to be understood as an instance of *type*, which comes before it. The signing in *type* is accompanied by a subtle raising of the eyebrows, a tilting of the head and a short transition between the two with no hold on the *type*. For example,

  instance-of(*glass, placements of cylinders*) (2)

  indicates that the cylindrical objects being placed on the table are glasses.

[2]There is also usually a difference between the classifier for a chair in LSF and ASL, but the signer in this LSF example actually uses a handshape very close to the ASL version, which is usually used in LSF for a small animal

(a) Place Plate (b) Place Two Glasses (c) Place Knife & Fork

Figure 2: Placements in the Table Description

- *place-object*(*loc*, *class*), is a rule that indicates the placement of an object in signing space. Here *class* is a classifier that indicates what kind of object is being placed at *loc*, which does not have to be one hand. The signer forms the classifier and with a downward settling motion with normal speed, places it at the *loc*. An example of this is

  place-object(*Midssp*, *prf-cylindrical-small*) (3)

  in which the signer's hand assumes C-handshape oriented vertically, figuratively around the cylinder, held just above the point, which here is "middle of signing space", and then settles the hand down to that point.

- *each-of*(*items*), here *items* is a list of things to be signed with emphasis on each individual element of the list. Each element is signed at a normal speed, the last posture of each element is held and a medium transition time is used between the items. An example of this is

  each-of(*two pairs of glasses*) (4)

  The signer places a pair of glasses simultaneously, one with each hand, then pauses before placing another pair.

- *all-of*(*items*), again *items* is a list of things to be signed, but in contrast with *each-of*, focus is placed on the group of elements from the list. Each element is signed at a faster than normal and the motion bounces at the end of each element instead of holding. An example of this is

  all-of(*four flat round plates*) (5)

  In this case, the signer uses both hands to show the round shape of the plate and a downward motion to place it, but instead of pausing, then proceeds to rapdily move with a bouncing motion to place the remaining three. The effect is seen as a group of plates as noted in (McDonald and Filhol, 2021).

One important aspect of AZee is that it does not presume a definition of "lexical" or "fixed" signs, even though classical linguistic models of signing would clearly indicate them in this discourse. For example, consider the signing GLASS before the classifier placements. In LSF, GLASS corresponds to an AZee rule defining that a C-handshape on the dominant hand is tapped twice vertically on the flat palm of the non-dominant hand. Since this is linked with an *instance-of* rule, the sign GLASS is performed while raising the eyebrows and chin, followed by the classifier placements, which situate the glasses in space. This same kind of pairing occurs in other sign languages as well. For example, in videos explaining ASL classifiers (ASL-That, 2012) (Handspeak, 1), the demonstrators display remarkably similar eyebrow and/or head motions after each lexical sign and before each classifier placement in mid signing space.

In this scene and in others that have been animated with the signing avatar Paula (Wolfe et al., 2011), the lowest common denominator among rules in AZee that seem to correspond to what linguistics would normally call a "lexical" sign is that it can be applied in the AZee description without any parameters at all, i.e. it can be signed generically. This cannot be said of any of the other rules listed above, all of which require parameters to be specified. For example, *place-object* must have both the classifier and location specified. There is no reasonable choice for a "default" object or "location" from AZee's perspective. This is different from the signs for RUG, TABLE, PLATE, etc., which can all be signed in a generic citation form. We will use the term "lexical" for AZee in this sense throughout this paper.

4. Exploiting common structures

The goal of synthesizing the same discourse in several languages can, of course, be seen as a translation task between sign languages, as in the present case study since the AZee description for the signing in question came from LSF. Our goal is then to take the description of signing in LSF that describes a setting of a table, and produce an equivalent understandable discourse in other languages. For this case study, we chose to work with DGS and GSL as the other target languages. The key is to look for structures in signing that they would seem to have in common.

Discussions with native signers, analysis of examples from corpora in both DGS and GSL, and the theoretical reasoning above concerning the strong similarities of placements and movements, reveal that classifier placements are likely to be done in a very similar manner across these sign languages. For the relative placement of object in a scene, while the specific classifier handshapes may differ in form, all three of these languages place objects using a downward settling motion. Since a large majority of the signing in discourse such as this description fall into this category, the current discourse is a good candidate for this case study, and we will begin with the following hypothesis:

Hypothesis 1 *AZee rules such as* place-object *and* move-object *have a similar form in each of these target languages.*

This hypothesis does, as we have noted, seem to have support in current linguistic theory. Note, however, that the geometrical presentation of its parameter *class* will generally not be the same, but we will assume that we can find a classifier for each object in question.

On the other end of the scale, lexical signs are often very different across sign languages, as can be seen with a casual inspection of sign dictionaries in any two sign languages. Even languages like LSF and ASL, which do share a common history and do have some cognates, are generally so different in their lexicons as to render many signs inscrutable to signers of the other language. From this, we can conclude that some mechanism will need to be used to animate the concept that corresponds to a lexical item in each sign language. Note that we will not assume here that it is possible in all cases to translate lexical signs in one language to single *lexical* signs in another, but the system will have to be flexible enough to handle this transfer.

Somewhere in the middle, we have all the processes that link these signs together, which we will call here *connecting rules*. In LSF, we have *instance-of*, which links each object's lexical sign to its corresponding placements, while whole sections of the discourse, such as the placement of the table as being the area where the rest of the setting is placed, is accomplished via *in-context* rules. Note that these were not arbitrarily chosen but rather the visual cues for these rules, such as eyebrow raises, head-tilts, pauses and blinks, were observed in the signer's movement. Here, there is a huge question concerning whether these kinds of linkages will be present in other languages.

As mentioned, something similar to the *instance-of* rule has been observed in ASL, and there are instances in other languages. For example, consider this example from the DGS Corpus, http://tinyurl.com/5chseh3u, where signer is signing "Das ganze Land, Deutschland", which begins with a shape specifier and the sign for LAND, accompanied by a similar raising of the head and eyebrows. The eyebrows are then lowered for the signing of DEUTSCH. This is similar to the motion described by *instance-of* in LSF. At least in this and several other instances, contextualization seems to be communicated in a similar way to LSF. Thus we will formulate a second hypothesis upon which this case study is based.

Hypothesis 2 *The connecting rules* instance-of*,* in-context*,* each-of*,* all-of *have a similar form in each of the target languages.*

Again, we are not claiming here that this AZee rule applies in general in DGS, but that it is perhaps reasonable to try synthesizing DGS discourse with this same AZee rules. Certainly, more study will be needed to confirm or disprove this. The trouble is that it has taken nearly ten years of corpus study to arrive at the current list of AZee rules for LSF. So, the question arises of whether there could be any transfer of learning that can shorten the time required to formulate AZee rules in other languages. This brings up an interesting possibility, which we formalize as a hypothesis that should be tested in the future.

Hypothesis 3 *If it is possible to synthesize sufficiently high quality animation using these descriptions, then the resulting videos could be tested with the Deaf community for their legibility and fidelity of the message. Furthermore, since the avatar system is generative and able to include or exclude features as necessary, individual linguistic features could be tested.*

When evaluating this hypothesis, it is important to note that such testing would be very difficult with motion capture or video of live signers because it is impossible for a human to reproduce a production precisely while including, omitting or changing only one linguistic parameter. Note also that the application of this last hypothesis is one way to test hypotheses 1 and 2

5. Avatar support

The goal of this case study is to accomplish the display of this table-description discourse in multiple languages, based on the same description in AZee. Following the discussion of the last section concerning AZee representations, and in particular Hypothesis 2, the current case study will use the connecting AZee rules *instance-of*, *in-context*, *each-of*, *all-of* and *place-object* as provided in the current AZee specification. Further study will show if changes to these rules will in general be necessary for a particular target language, or whether they need to be replaced by other connecting rules in a particular target.

Avatar synthesis of the signing will be accomplished using the previously published AZee-Paula bridge (McDonald and Filhol, 2021). This prior work also details how this table scene was synthesized in LSF. To extend this bridge to support animation in the target languages, there are two main classes of rules that need to be transformed into the target languages, GSL and DGS:

1. The seven lexical items, *rug*, *table*, *chair*, *plate*, *glass*, *knife*, *fork* which in this scene are signed essentially in their citation form.

2. Classifier specifications for *class-generic-object*[3], *class-flat-round-large*, *class-cylindrical-small*, *class-straight-elongated* and *class-large-rectangle*.

For the lexical items, recall that the Azee-Paula bridge contains a system of shortcuts which map such rules to pre-animated sequences on Paula which can then be blended with the motions defined by the connecting rules. In the target languages chosen, fixed signs exist to encompass these concepts and so the transformation is as simple as mapping these rules to the corresponding glosses in the target language. It is important to note that this may not always be the case, and a more complicated transfer may be necessary. This does not occur for the chosen target languages in this table description, and extensions to this mechanism will be explored in the future.

For the classifiers, recall that the bridge uses a system of artist exemplars that set up the configuration of the hand and arm which provides an example posture to build the classifier placement or motion from. The system then uses inverse kinematics to accomplish the placement or movement while using the data from the artist template to inform choices for redundant degrees of freedom (McDonald and Filhol, 2021). Again, all that is necessary is that Paula's database in the target language provides an artist shortcut for each of the classifiers used in the discourse. These may be very different from the ones provided for LSF, and may eventually involve changes to the AZee rules themselves in the target language. However, this does not occur in this case study's table scene description.

[3]The signer in LSF used a non-standard classifier for chair which is usually used instead for generic objects, but which is incidentally very close to the classifier used in ASL, DGS and GSL for a chair or a sitting person

6. Results

The main database for a language in the Paula system contains all of the information that the avatar needs to animate sign directly from the linguistic description, including pre-animated citation forms of lexical signs and classifier definitions. The main task on the avatar side of this case study was two-fold

- Animate each of the seven lexical items in DGS and GSL in their corresponding databases.

- Set the AZee shortcuts for the rules corresponding to these lexical items to link to these new animations in each database.

Both of these steps were completed for all three languages so that the correct animation in each language was automatically triggered as a shortcut to the corresponding AZee rule. Figures 3 and 4 contain still frames from two of the seven lexical items in the three languages.

The other main difference that can certainly exist between sign languages is the form of classifiers, and thus it is in general necessary to create artist exemplars of each classifier. As part of this initial study, we showed the set of classifier forms to a small group of sign language experts from both the University of Hamburg and the Athena Research Center in Athens. Both groups indicated that the classifiers that were used in the LSF discourse were acceptable for their respective sign languages. Thus the original classifier exemplars built for LSF were transferred to the GSL and DGS databases to perform the synthesis. With these changes, the Paula system could synthesize the table scene description directly and automatically from the original AZee description. The results are displayed in the accompanying video which can be accessed at http://tinyurl.com/2ch2bwwm.

7. Conclusion and future work

This paper presents a first step towards a multilingual display system in the form of a case study, and as such is limited in scope, but nevertheless points the way towards further development. As noted above, there are several avenues that must be investigated:

1. For this line of research to proceed, the resulting synthesized sign must be tested with the Deaf community. Testing with fluent signers is essential for both formative and evaluative feedback on the system, and will be a key element of testing the three hypotheses outlined in section 4.

(a) LSF, tapis (b) DGS, teppich (c) GSL, χαλι

Figure 3: Images from the animations for RUG in the three languages

(a) LSF, chaise (b) DGS, stuhl (c) GSL, καθισμα

Figure 4: Images from the animations for CHAIR in the three languages

2. One of the main roadblocks is vocabulary acquisition in each of the target languages, however, if signing of sufficiently high quality could be derived from HamNoSys or another phonetic description, existing corpora could address this lack.
3. Classifiers may differ significantly, not only in handshape, but also in the orientation of the hand during placement and movement. An example of this is the classifiers for vehicles in ASL and LSF. Paula can handle part of this difference with the artist exemplar, but changes to the AZee rule for the classifier will also be necessary.
4. The forms of the linking rules described here may differ in other languages, and completely different linking rules may be discovered that don't translate between languages.

Pursuing each of these, with repeated user testing with the Deaf community will point the way to extending this effort to more general multilingual display of sign language.

Acknowledgement

This work is supported in part by the EASIER (Intelligent Automatic Sign Language Translation) Project. EASIER has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement n° 101016982.

8. Bibliographical References

References

Nicoletta Adamo-Villani and Ronnie B Wilbur. 2015. ASL-pro: American sign language animation with prosodic elements. In *Universal Access in Human-Computer Interaction. Access to Interaction: 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II 9*, pages 307–318. Springer.

Akorbi. 2023. Is sign language universal? https://akorbi.com/blog/is-sign-language-universal/, Accessed: 2024-02-05.

ASL-That. 2012. ASL Classifiers (CLs) for Furniture & Objects. https://www.youtube.com/watch?v=xPb8AD1rON0, Accessed: 2024-02-11.

Mohamed-El-Fatah Benchiheub, Bastien Berret, and Annelies Braffort. 2016. Collecting and analysing a motion-capture corpus of french sign language. In *sign-lang@ LREC 2016*, pages 7–12. European Language Resources Association (ELRA).

Carl Börstell. 2023. Ableist Language Teching over Sign Language Research. In *Proceedings of* *the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)*, pages 1–10.

George Caridakis, Stylianos Asteriadis, and Kostas Karpouzis. 2011. Non-manual cues in automatic sign language recognition. In *Proceedings of the 4th international conference on pervasive technologies related to assistive environments*, pages 1–4.

Shatabdi Choudhury. 2022. Analysis of Torso Movement for Signing Avatar Using Deep Learning. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives*, pages 7–12.

Deutscher Gehörlosen-Bund et. al. 2024. ”sofortiger handlungsbedarf” – deutsche gehörlosendachverbände gegen avatare. https://www.taubenschlag.de/2024/02/sofortiger-handlungsbedarf-deutsche-gehoerlosen-dachverbaende-gegen-avatare/, Accessed: 2024-02-23.

Dictasign. 2012. Sign language recognition, generation and modelling with application in deaf communication. https://cordis.europa.eu/project/id/231135, Accessed: 2024-02-11.

EASIER-Project. 2024. Intelligent automatic sign language translation. DOI:https://doi.org/10.3030/101016982, Accessed: 2024-02-11.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2022. *Ethnologue: Languages of the World*, 25 edition. SIL International, Dallas.

Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Goudenove. 2010. DICTA-SIGN: sign language recognition, generation and modelling with application in deaf communication. In *sign-lang@ LREC 2010*, pages 80–83. European Language Resources Association (ELRA).

Michael Filhol, Mohamed Nassime Hadjadj, and Annick Choisier. 2014. Non-manual features: the right to indifference. In *sign-lang@ LREC 2014*, pages 49–54. European Language Resources Association (ELRA).

Michael Filhol, John McDonald, and Rosalee Wolfe. 2017. Synthesizing sign language by connecting linguistically structured descriptions to a multi-track animation system. In *International Conference on Universal Access in Human-Computer Interaction*, pages 27–40. Springer.

Michael Filhol, Thomas von Aschenberg, and John McDonald. 2024. Final Integration Allowing Post-Editing with Discourse Representation Diagrams. https://www.project-easier.eu/deliverables/.

Susan D Fischer. 2015. Sign languages in their historical context. In *The Routledge handbook of historical linguistics*, pages 442–465. Routledge.

Sylvie Gibet and Pierre-François Marteau. 2023. A Text-To-SL Synthesis System Using 3D Avatar Technology. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5. IEEE.

Handspeak. 1. ASL Classifiers: [objects] on a table. https://www.youtube.com/watch?v=GF0iqbD64TA, Accessed: 2024-02-11.

Handspeak. 2. "onomatopoeia" in sign language. https://www.handspeak.com/word/4056/, Accessed: 2024-02-11.

Thomas Hanke. 2004. HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC*, volume 4, pages 1–6.

Thomas Hanke, Lutz König, Reiner Konrad, Maria Kopf, Marc Schulder, and Rosalee Wolfe. 2023. EASIER Notation–a proposal for a gloss-based scripting language for sign language generation based on lexical data. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5. IEEE.

Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in size and depth. In *sign-lang@ LREC 2020*, pages 75–82. European Language Resources Association (ELRA).

Annette Hohenberger. 2007. The possible range of variation between sign languages: Universal grammar, modality, and typological aspects. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, 188:341.

Michael Kipp, Alexis Heloir, and Quan Nguyen. 2011. Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*, pages 113–126. Springer.

LIMSI, CIAMS, and LISN. 2022. MOCAP1. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

John McDonald and Michael Filhol. 2021. Natural synthesis of productive forms from structured descriptions of sign language. *Machine Translation*, 35(3):363–386.

Rachel Locker McKee and Jemina Napier. 2002. Interpreting into international sign pidgin: An analysis. *Sign language & linguistics*, 5(1):27–54.

Johanna Mesch. 2010. Perspectives on the concept and definition of International Sign. *World Federation of the Deaf*.

Marcus Perlman, Hannah Little, Bill Thompson, and Robin L Thompson. 2018. Iconicity in signed and spoken vocabulary: a comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in psychology*, 9:1433.

Roland Pfau, Markus Steinbach, and Bencie Woll. 2012. *Sign language: An international handbook*, volume 37. Walter de Gruyter.

Oliver Sacks. 2022. *Seeing voices: A journey into the world of the deaf*. Neha Publishers & Distributors.

Wendy Sandler. 2010. Prosody and syntax in sign languages. *Transactions of the philological society*, 108(3):298–328.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2021. Continuous 3d multi-channel sign language production via progressive transformers and mixture density networks. *International journal of computer vision*, 129(7):2113–2135.

Rosalee Wolfe, John Mcdonald, and Jerry Schnepp. 2011. An Avatar to Depict Sign Language: Building from Reusable Hand Animation. In *International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, Berlin, Germany.

Rosalee Wolfe, John C McDonald, Thomas Hanke, Sarah Ebling, Davy Van Landuyt, Frankie Picron, Verena Krausneker, Eleni Efthimiou, Evita Fotinea, and Annelies Braffort. 2022. Sign language avatars: A question of representation. *Information*, 13(4):206.

Inge Zwitserlood. 2012. Classifiers. In *Sign language: An international handbook*. De Gruyter.