A Text-To-SL Synthesis System Using 3D Avatar Technology

Sylvie Gibet, Pierre-François Marteau

A TEXT-TO-SL SYNTHESIS SYSTEM USING 3D AVATAR TECHNOLOGY

Sylvie Gibet, Pierre-François Marteau

IRISA, University of Bretagne Sud, F-56000 Vannes, France
sylvie.gibet@irisa.fr

ABSTRACT

In this paper, we present our Text-to-SL system that synthesizes sign language (SL) from a textual specification by means of a 3D virtual signer. This system is based on a multi-channel composition mechanism, which spatially and temporally arranges linguistic elements to generate SL utterances. New hybrid synthesis modules are proposed that enrich our initial *SignCom* system, combining procedural and data-driven synthesis. They implement grammatical dynamics specific to SL, in particular grammatical inflection mechanisms based on the fundamental concepts of spatiality and iconicity. We also present a methodology and perspectives for automatically translating written text in a spoken language into sign language.

Index Terms— sign language, synthesis, translation, signing avatar

1. INTRODUCTION

Deaf people still face difficulties in communicating in their sign language (SL). In order to facilitate access to information, both in SL and in spoken language (oral or written), we are developing digital tools for deaf signers that automatically produce SL from written text. Currently, most of the available applications for the Deaf are based on video. However, if video is the most shared media by deaf people, it does not guarantee anonymity and imposes strong constraints of storage and transport of information. On the other hand, the automatic production of SL and visualization using 3D avatars is an alternative answer that makes it possible to reduce stored information, anonymize and manipulate it to edit, view, and produce new utterances at a lower cost.

In this paper, we focus on Text-to-SL systems that generate, synthesize and translate SL from text, using 3D signing avatars; we will group them under the acronym TSL. We present our system *SignCom* which is an extended version of the system presented in [1]. This system is based on a compositional principle for constructing and editing signs and utterances in LSF, by combining motion data from the phonological to the syntactic/semantic levels. Before evoking the challenges posed by the new models of artificial intelligence, we describe more specifically our technological advances made in recent years, which concern hybrid synthesis, coupling data-based and procedural synthesis.

2. RELATED WORK

Studies on TSL systems are relatively recent. In this section, we review the main approaches related to linguistic representations of SL and TSL systems, separating Text-to-SL synthesis systems and automatic translation systems; for an exhaustive review on the subject, refer to [2] and [3].

Linguistic representations of SL. Early work on SL phonology led to different types of representations. Among these, the work of [4] led to the description of ASL as a combination of elementary units constituting signs: the location of the sign, the shape of the hand and its movement. One of the basic assumptions is that two distinct signs can be differentiated when only one of the constituent parameters is changed (notion of minimal pairs). Continuing Stokoe's work, other parameters that participate in the formation and distinction of signs have been identified, including hand orientation and non-manual elements such as facial expressions, eye gaze, torso orientations, and certain body gestures [5]. These phonological elements are arranged and synchronized spatially and temporally to form signs and utterances in SL. The notation system *HamNoSys* [6] takes the previous parameters and transcribes the signs in a linear way using the computer symbols *Unicode*. The remarkable linguistic work on ASL by Liddell and Johnson has led to the definition of a phonetic model based on the Posture-Detention-Transition-Shift (PDTS) scheme, distinguishing on each phonological component – hand configuration (HC), orientation (FA), placement (PL), non-manual features (NM) – static elements and dynamic transitional elements [7].

In *computational linguistics*, SL gestures have been described using formalisms ranging from scripts to dedicated computer languages. Thus, the language *SiGML* [8] based on *HamNoSys* has been developed to generate 3D avatar animations. This language was then extended by incorporating Johnson's PDTS model [9]. More recently, the formal language *Azee*, based on abstract modeling, integrates a geometric formalism and aims at representing a set of grammatical processes in a parallelized and non-linear way [10]. These scripting or specification languages are able to describe signs or utterances in a very analytical and precise way. However, the specification of new signs can be very tedious. Moreover, most of these languages integrate explicit temporal elements into their formalism. This is notably the case of *SIGML*, *EMBRscript* [11] and *Azee*, for which the key postures of the avatar are specified at pre-determined moments. In contrast, the partition/constitution (P/C) model proposes an implicit synchronization scheme based on a 2D representation of a 3D syntactic graph. This latter model facilitates the joint visualization of the temporal and spatial axis (channels), while managing the coordination of the different channels.

Several linguistic representations have focused on grammatical inflections of signs. For example, the Italian sign language project *ATLAS* [12] incorporates inflectional processes involving placement, hand configuration and movement, as well as size and shape specifiers. For its part, the *AZee-Paula* [13] system allows the generation of *proforms* as well as several inflectional mechanisms in ASL.

Text-to-SL animation systems. Among the technologies available to animate signing avatars, there are three main approaches.

The first one consists, from restricted data, to animate an avatar using so-called *pure synthesis* methods. We group in this category

the approaches based on key postures, either manually or automatically determined, associated with interpolation processes, and procedural approaches that automate the movement generation process. While these systems, which couple a specification language to an animation engine, are capable of achieving goals of precision and fine temporal control, they generally result in unnatural, even robotic movements. Moreover, due to a long and tedious specification, they lead to the creation of a limited number of signs, with little or no possibility of grammatical inflections. For example, the *EMBR* [14] system is used to generate standard signs from sequences of keyposes specified using the *EMBRscript* language. The animation engine *AnimGen* is integrated with *JASigning* to create signs specified with *SiGML* [8,15]. These animation systems have been used for different SL [16–18]. In both cases, sign inflection is only possible at the lexical level. More recently, the *Paula* system developed at DePaul University has proven capable of generating animations from the *AZee* representation, combining techniques based on key postures and procedural algorithms that improve the fluidity of movements and facilitate multimodal synthesis [19, 20].

The second approach, called *data-based synthesis*, reuses the movement as a basic resource. In this case, motion is captured by motion capture (MoCap) technology that simultaneously records hand, body, and facial expressions, as well as eye gaze. The projects *SignCom* [1], *Sign3D* [21] or *Rosetta* [22] exploit these synthesis principles, based on data captured at high resolution.

The two types of control (data-based and hand-crafted synthesis) can be combined, to become the so-called hybrid synthesis. It is indeed possible to replace synthesis from scratch by motion data previously recorded and annotated, or combine procedural methods with data through machine learning algorithms. This hybrid control gives some flexibility and variational possibilities in the generated signed sequences. These approaches have been developed in several research teams: coupling data-driven and machine learning for directional verbs study [23], coupling hand-crafted and kinematics methods *SLPA/Azee* [10, 19], or coupling data-driven and inverse kinematics for spatial inflecting signs [24].

Automatic translation systems. With the advent of deep learning, recent approaches in Neural Machine Translation (NMT) have been successfully developed in SL, inspired by text-to-speech translation or from one spoken language to another. Several review papers have been presented on text-to-SL and SL-to-text [3, 25, 26]. In this context, NMT have been proposed for Arabic SL [27], or for translating spoken English into ASL. The system proposed by [28] uses GANs associated with motion graphs to produce SL videos from spoken language sentences. Video-based corpora have been created to support NMT work, mainly in recording studios with one or more cameras. An overview of European corpora is presented in [29]. The *RWTH-PHOENIX-corpus*, used in many studies of video sign recognition, includes 1980 German Sign Language (GSL) sentences describing weather forecasts [30]. Recently, the *Rosetta* project has led to the creation of the Rosetta-LSF [31] corpus which contains 3 hours of LSF in a journalistic domain.

3. OUR TEXT-TO-SL SIGNCOM SYSTEM.

Sign languages require a high level of precision and realism in the execution of body, hand and facial movements in order to be understood and accepted by the Deaf. The animation of a 3D avatar based on motion capture (MoCap) makes it possible to meet these requirements. This led us to develop our concatenative synthesis system *SignCom* that simultaneously produces hand and body movements, facial expressions and eye gaze. This system facilitates the construction of new sentences, cutting/pasting/mixing/interpolating selected motion segments, and recombining them to automatically produce the 3D animation of a signing avatar [1]. After recalling the principle of concatenative synthesis, we describe below the different synthesis modules that enrich our initial system, coupling procedural hand-arm synthesis and data-driven synthesis, while preserving grammatical inflection mechanisms.

3.1. Principle of Concatenative synthesis

Our motion data are recorded thanks to a MoCap system based on passive markers and infrared cameras, which measures very precisely the position of the markers placed on the body (42 markers), the hands (22 markers per hand), and the actor's face (41 markers) with a frame rate of 200 Hz. It results from the recording session a dataset of motion sequences representing the utterances of the corpus in LSF. These motion data are then annotated following multi-channel schemes for which we distinguish i) linguistic channels that respect the phonological description of the signs [7] and their grammatical class [32], and ii) physical channels that correspond to groups of articulations (hand, arm, etc.).

The synthesis system is divided into two parts: an off-line process of data storage and an on-line process of data extraction and animation control. The need to encode in the signed statements, on the one hand the linguistic information translating the multi-linear structure of the LS and on the other hand the motion flows, requires to build beforehand two coupled heterogeneous databases, one containing the semantic data, the other the motion data (positions or angles of the skeleton joints). The semantic database performs the coupling between symbolic data following our multi-channel annotation scheme and a list of movements indexed by means of a name, temporal markers, and a sequence of articulations involved in the motion. This is a one-to-many coupling to account for the existence of several instances of the same sign or part of a sign in the corpus. A multi-condition query language allows, from the specification of elements of this annotation, to automatically extract a continuous stream of movement postures that correspond to it.

Different editing processes have been explored with this system, i) by replacing signs or groups of signs, ii) by instantiating predefined syntactic patterns, or, iii) by replacing phonological components of the signs – hand configurations, hand movements, torso or head movements, and facial expressions. However, while the first version of the system was able to manipulate SL content at the phonological, lexical, or discursive levels, based on manual annotation, it was not able to synthesize more complex grammatical processes, including spatiality and iconicity. We therefore developed an extension of the *SignCom* system that incorporates some inflectional processes of the LSF grammar [33]. Figure 1 illustrates this system as a whole, integrating the basic functional modules (in blue) and those constituting the technological advances (in pink).

3.2. Enrichment of the dataset by hybrid synthesis

The aim of body synthesis, known as *hybrid*, is to enrich the initial database by adding to it data synthesized with inflectional variations, thus facilitating the creation of new utterances in LSF [24]. Indeed, capturing large amounts of movement is time-consuming and tedious, and there are still few MoCap corpora covering the large variability of SL. We have developed synthesis techniques that increase the recorded data while respecting some mechanisms of spatiality and iconicity of SL. In particular, to add variability to the data, one possibility is to use the Local Grid Bootstrap algorithm to gen-

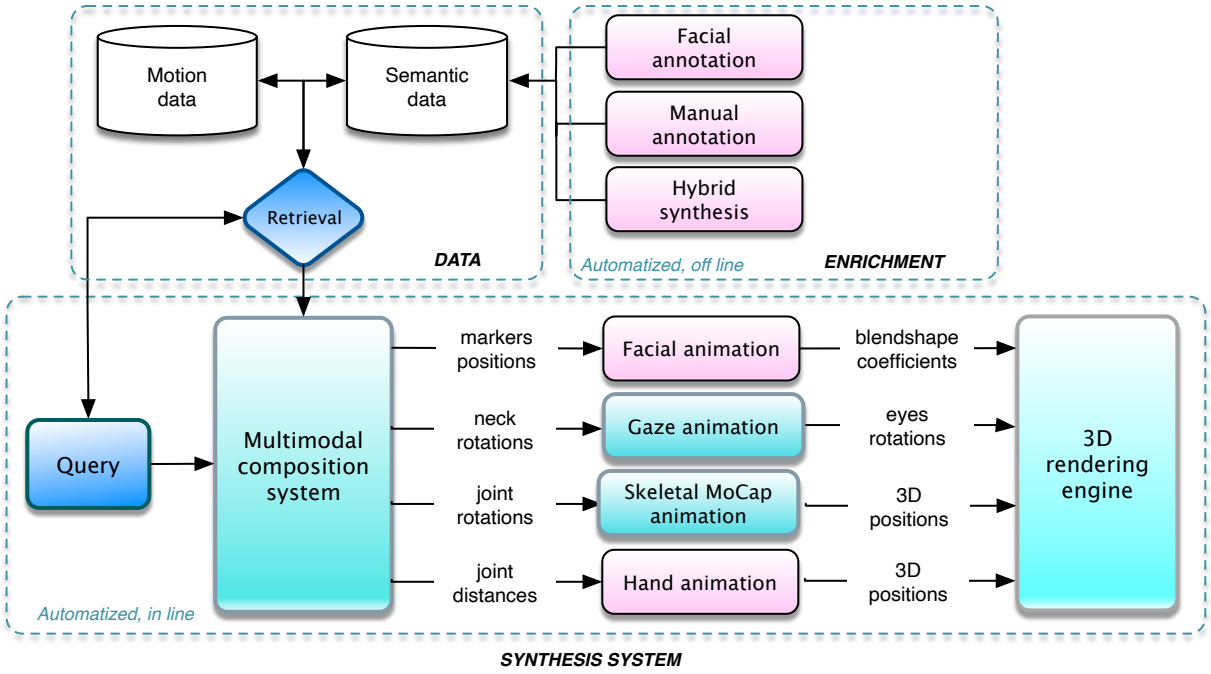

Fig. 1. Text-to-SL *SignCom* synthesis system with signing avatar. In blue: basic system [1]. In pink: advanced technological modules

erate new expressive trajectories of the end effector [34]. Some of these techniques have been implemented and evaluated (pointing, spatialization, lexical and syntactic modifications from HC manipulation) [24]; other processes are specified in this paper (indicating verbs, size and shape specifiers) for integration into our *SignCom* system.

Pointing gestures use loci (targets) located in discretized spatial areas and pointed movements toward these loci. Such gestures lead to syntactic processes, as they may involve the referencing of discourse entities located in the signing space, or have a pronominal function. However, by synthesis, it is possible to produce an unlimited number of pointing gestures that cover the signing space. To this end, we have built efficient kinematic inversion models (IK) which, from the sole specification of a set of referenced target positions (loci) [34], generate automatically the pointing motion towards these targets [24]. We propose to introduce geometric constraints that enable to preserve the grammatical consistency of SL [35].

Spatialization. The same process can be used to address the issue of sign spatialization. Indeed, having the signs executed in the area of their lexical anchorage, with the IK technique it is possible to move the hand towards the desired area and to sign the lexical entity at this specific location.

Indicating verbs are characterized by a trajectory that defines the path of motion from one locus to another and thus make possible the distribution of the actancial roles agent/beneficiary/object in the sentence. By specifying the loci corresponding to the desired pronouns, and then by applying IK to these loci, it becomes possible to synthesize the hand movement. For example, the sentence *I give you a book*, can be translated into LSF by the ID-glosses parameterized function: [TO GIVE]([PRO-1],[PRO-2],[BOOK]) and signed by a movement of the hand, from the location of the pronoun *I* [PRO-1] to the location of the pronoun *You* [PRO-2], with a manual configuration (proform) which is that of the sign [BOOK]. To generate the sentence *You give me a book*, only the movement from [PRO-2] to [PRO-1] needs to be reversed. To generate the sentence *I give him a book*, the new locus corresponding to the 3rd person [PRO-3] is specified, and the movement is synthesized by IK from [PRO-1] to [PRO-3] while keeping the hand configuration. For indicating verbs with a direct object, the sentence in LSF is signed with a HC that corresponds to this object. This synthesis process can be generalized to most sentences with indicating verbs.

Size and shape specifiers (SSS). We distinguish SSS that act on movements, hand configurations or hand orientations, or on a combination of these components of signs, and we propose below a non-exhaustive list of sign inflections encountered in LSF. These SSS have been modeled and specified. They involve different synthesis mechanisms.

- The size of the motion path can be changed by specifying a straight path in the cross-sectional plane of the sign and synthesizing the motion using an IK or interpolation technique. This is the case with the [TABLE] sign, in which the flat hands move more or less apart depending on the size of the table.
- It is relatively easy to specify and replace the static HC on the whole sign, as for example in the signs [GLASS-END], [LARGE-GLASS], where only the shape of the hand changes (more or less open) [24].
- Dynamically changing the HC during a sign requires the use of a direct kinematics (DK) model between two or more key hand shapes. For example, from the [GLASS] sign, it is possible to generate the [CUP-OF-CHAMPAGNE] sign, in which

the hand shape flares upwards.

- Other SSS involve changing the trajectory and orientation of the hand simultaneously, with the HC remaining static. This is the case for the sign [BOWL], which can vary in size. Similarly, trajectories and manual configurations can be modified simultaneously, the orientation being static, as in the sign [BANANA]. Synthesis techniques consist of sketching and generating more or less complex trajectories, synthesizing by IK the movement of the arm joints, and synthesizing by FK the dynamic HC or orientations.

Evaluation of the synthesis system. Several perceptual evaluations were carried out for the enrichment of the synthesis.

In [34], we tested the generation of motion through IK process. We showed that end-effectors trajectories provided an efficient way to control virtual characters while preserving the main expressive cues contained in the original motion.

In [24], we conducted two perceptual evaluations on LSF synthesis via web surveys. The first one concerned the study of spatialization and pointing. 57 participants with very good LSF skills were asked to answer questions presented in the form of text (drop-down menus), images or videos. For the recognition task of the sign's location [BOWL] through the viewing of 8 videos (5 synthesized videos and 3 replayed videos), the recognition rate was 86% for the synthesized signs versus 63% for the replayed signs, the difference being explained by a greater variability of the real signs. Assessment of the realism of these same signs gave a mean score of 3.6 for the synthesized utterances versus 3.8 for the playback ones, on a 5-point Likert scale, showing that there was no significant difference between synthesized and replayed sequences (*p-value* of 0.031). Concerning the assessment of the realism of the pointing gestures, 9 videos were presented to the same participants (6 synthesized and 3 replayed videos). The scores were 3.15 for the synthesized gestures and 3.45 for the replayed gestures, respectively. Again, we concluded that there was no marked difference between the real and synthetic gestures (*p-value* of 0.081). The second evaluation concerned processes related to the manipulation of static or dynamic manual configurations (DHC). For this study, 39 participants viewed 20 videos presented in a random order (13 synthesized and 7 replayed), representing signs from dactylology such as [LSF] or [OK], or different animal gaits (e.g., rooster and cat). For the task of evaluating the replacement of DHC by other DHC (5 videos), the results showed that there was no significant difference between replayed and synthesized utterances, the latter being considered even more realistic. For the task of recognizing dactylological signs (15 videos), the recognition rates obtained for synthesis were 95% for the synthesized signs and 91% for the replayed signs. Furthermore, no significant difference was observed in the realism of the synthesized (score of 3.08/5) and replayed (score of 3.03/5) signs.

4. TEXT-TO-SL TRANSLATION: METHODOLOGY AND PERSPECTIVES

In the near future, we would like to extend this work by modeling most of the grammatical mechanisms found in SL and by automating the translation process from spoken/written languages (WL) to sign languages (SL).

Translating Text-to-SL can be performed in two steps: 1) a *Text to Pivot-SL generation* step involving the translation of the written text in a spoken language – possibly generated from audio signal – into the computational representation of SL (so-called *Pivot-SL*); 2) a *Pivot-SL to 3D motion* step representing the translation of the symbolic SL representation into the multi-channel specification that produces a continuous 3D motion stream, thanks to the animation engine. The two steps are approached differently. For the first step, it is possible to use rule-based algorithms to model the two WL and SL languages, thanks to ontologies (semantic network) including the identified syntactic concepts, or to use a pivot language, intermediate between spoken language in its written form (WL) and glossed-SL, such as *AZee* [36], in order to automatically generate a transcription of SL and thus compensate for the lack of data. The second step is based on the specification of a sentence in glossed-SL to automatically produce the animation of the signing avatar in 3D, thanks to the *SignCom* animation system. An alternative approach is to propose recent deep learning *end-to-end* methods [28, 37] to automatically translate written WL to SL, using for example *transformers*.

While the advances in Text-to-SL systems with signing avatars are very promising, many research questions remain open. In particular, data available for training neural network models remains limited, and more generally existing video corpora may be insufficient to cover the large variability of SL due to the many grammatical inflections. Moreover, compared to spoken languages that encode linearly both the written word – as a sequence of phonetic elements – and their corresponding sound units, there is a lack of aligned resources between text and SL required to provide parallel resources for training models. To overcome this lack of labeled motion data, the solution could be to go through an intermediate language that would guide the learning process by introducing linguistic knowledge. However, the lack of a commonly accepted written representation for SL does not facilitate the recording of sufficient information to match a text in an oral language to its transcription in SL.

Nevertheless, the possibility of capturing large volumes of data and the development of deep machine learning methods will open up prospects for automatic translation systems from spoken languages to SL or vice versa, or from one SL to another.

5. REFERENCES

[1] S. Gibet, N. Courty, K. Duarte, and T. Le Naour, "The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation," in *ACM Trans. on Interactive Intelligent Systems*, 2011, vol. 1:1.

[2] Lucie Naert, Caroline Larboulette, and Sylvie Gibet, "A survey on the animation of signing avatars: From sign representation to utterance synthesis," *Comput. Graph.*, vol. 92, pp. 76–98, 2020.

[3] Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka, "A survey on sign language machine translation," *Expert Systems with Applications*, vol. 213, pp. 118993, 2023.

[4] William C. Stokoe, *Semiotics and Human Sign Language*, Walter de Gruyter Inc., 1972.

[5] Robbin Battison, *Lexical borrowing in American sign language.*, ERIC, 1978.

[6] Siegmund Prillwitz and Hamburg Zentrum, *HamNoSys: Version 2.0; Hamburg Notation System for Sign Languages; An Introductory Guide*, Signum-Verlag, 1989.

[7] R.E. Johnson and S.K. Liddell, "A segmental framework for representing signs phonetically," *Sign Language Studies*, vol. 11, no. 3, pp. 408–463, 2011.

[8] R. Elliott, J. R.W. Glauert, JR Kennaway, I. Marshall, and E. Safar, "Linguistic modelling and language-processing tech-

nologies for avatar-based sign language presentation," *Universal Access in the Information Society*, vol. 6, no. 4, pp. 375–391, 2008.

[9] J. Glauert and R. Elliott, "Extending the sigml notation: a progress report," in *Int. Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, 2011, vol. 23.

[10] M. Filhol, J. McDonald, and R. Wolfe, "Synthesizing sign language by connecting linguistically structured descriptions to a multi-track animation system," in *Int. Conf. on Universal Access in Human-Computer Interaction*. Springer, 2017, pp. 27–40.

[11] A. Héloir and M. Kipp, "Real-time animation of interactive agents: Specification and realization," *Applied Artificial Intelligence*, vol. 24, no. 6, pp. 510–529, 2010.

[12] V. Lombardo, F. Nunnari, and R. Damiano, "A virtual interpreter for the italian sign language," in *International Conference on Intelligent Virtual Agents*. Springer, 2010, pp. 201–207.

[13] M. Filhol and J. Mcdonald, "Extending the AZee-Paula shortcuts to enable natural proform synthesis," in *Workshop on the Representation and Processing of Sign Languages, LREC 2018*, Japan, 2018.

[14] M. Kipp, A. Héloir, and Q. Nguyen, "Sign language avatars: Animation and comprehensibility," in *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, 2011.

[15] Richard Kennaway, John RW Glauert, and Inge Zwitserlood, "Providing signed content on the internet by synthesized animation," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 3, pp. 15, 2007.

[16] Sarah Ebling, John RW Glauert, JR Kennaway, Ian Marshall, and Eva Safar, "Building a swiss german sign language avatar with jasigning and evaluating it among the deaf community," *Universal Access in the Information Society*, vol. 15, pp. 577–587, 2016.

[17] E. Efthimiou, S.E. Fontinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Goudenove, "Dictasign–sign language recognition, generation and modelling: a research effort with applications in deaf communication," in *Workshop on the Representation and Processing of Sign Languages, LREC 2010*, 2010, pp. 80–83.

[18] F. Roelofsen, L. Esselink, S. Mende-Gillings, and A Smeijers, "Sign language translation in a healthcare setting," in *In Translation and Interpreting Technology Online (TRITON)*, 2021, pp. 110–124.

[19] J. McDonald, R. Wolfe, J. Schnepp, J. Hochgesang, D. Jamrozik, M. Stumbo, L. Berke, M. Bialek, and F. Thomas, "An automated technique for real-time production of lifelike animations of american sign language," *Universal Access in the Information Society*, vol. 15, no. 4, pp. 551–566, 2016.

[20] J. McDonald and M. Filhol, "Natural synthesis of productive forms from structured descriptions of sign language," *Machine Translation*, vol. 35, no. 3, pp. 363–386, 2021.

[21] Sylvie Gibet, François Lefebvre-Albaret, Ludovic Hamon, Rémi Brun, and Ahmed Turki, "Interactive editing in french sign language dedicated to virtual signers: requirements and challenges," *Universal Access in the Information Society*, vol. 15, no. 4, pp. 525–539, 2016.

[22] B. Dauriac, A. Braffort, and E. Bertin-Lemée, "Example-based multilinear sign language generation from a hierarchical representation," in *Sign Language Translation and Avatar Technology; Junction of the Visual and the Textual; Challenges and Perspectives*, 2022, pp. 21–28.

[23] Matt Huenerfauth, Pengfei Lu, and Hernisa Kacorri, "Synthesizing and evaluating animations of american sign language verbs modeled from motion-capture data," in *SLPAT@Interspeech*, 2015.

[24] L. Naert, C. Larboulette, and S. Gibet, "Motion synthesis and editing for the generation of new sign language content," *Machine Translation*, vol. 35, no. 3, pp. 405–430, 2021.

[25] Navroz Kahlon and Williamjeet Singh, "Machine translation from text to sign language: a systematic review," *Universal Access in the Information Society*, 07 2021.

[26] U. Farooq, M. Rahim, N. Sabir, A. Hussain, and A. Abid, "Advances in machine translation for sign language: approaches, limitations, and challenges," *Neural Computing and Applications*, vol. 33, 11 2021.

[27] Mourad Brour and Abderrahim Benabbou, "Atlaslang mts 1: Arabic text language into arabic sign language machine translation system," *Proc. Computer Science*, vol. 148, pp. 236–245, 2019.

[28] Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden, "Text2sign: towards sign language production using neural machine translation and generative adversarial networks," *International Journal of Computer Vision*, vol. 128, pp. 891–908, 2020.

[29] Maria Kopf, Marc Schulder, and Thomas Hanke, "Overview of Datasets for the Sign Languages of Europe," 2021.

[30] J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. Piater, and H. Ney, "Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus," 2012.

[31] LIMSI and LISN, "rosetta-lsf," 2022, ORTOLANG (Open Resources and TOols for LANGuage), www.ortolang.fr.

[32] Trevor Johnston, "The lexical database of AUSLAN (Australian Sign Language)," in *Proceedings of the First Intersign Workshop: Lexical Databases*, Hamburg, 1998.

[33] Agnès Millet, *Grammaire descriptive de la langue des signes française: dynamiques iconiques et linguistique générale*, UGA Editions, 2019.

[34] P. Carreno-Medrano, S. Gibet, and P.F. Marteau, "Perceptual validation for the generation of expressive movements from end-effector trajectories," *ACM Trans. Interact. Intell. Syst.*, vol. 8:3, jul 2018.

[35] Thibaut Le Naour, Nicolas Courty, and Sylvie Gibet, "Kinematics in the metric space," *Computers Graphics*, vol. 84, pp. 13–23, 2019.

[36] C. Challant and M. Filhol, "A first corpus of AZee discourse expressions," in *Proc. of the LREC 2022 Conference, France*, Marseille, France, 2022, pp. 1560–1565.

[37] D. Metaxas, M. Dilsizian, and C. Neidle, "Scalable ASL sign recognition using model-based machine learning and linguistically annotated corpora," in *Workshop on the Representation and Processing of Sign Languages, LREC 2018*, Miyazaki, Japan, 2018, pp. 127–132.