Journal of Universal Computer Science, vol. 17, no. 3 (2011), 399-425
submitted: 3/8/10, accepted: 28/1/11, appeared: 1/2/11 © J.UCS

The Synthesis of LSE Classifiers: From Representation to Evaluation

Fernando López-Colino
(Universidad Autónoma de Madrid, Spain
Dpt. de Tecnología Electrónica y de las Telecomunicaciones
fj.lopez@uam.es)

José Colás
(Universidad Autónoma de Madrid, Spain
Dpt. de Tecnología Electrónica y de las Telecomunicaciones
jose.colas@uam.es)

Abstract: This work presents a first approach to the synthesis of Spanish Sign Language's (LSE) *Classifier Constructions* (CCs). All current attempts at the automatic synthesis of LSE simply create the animations corresponding to sequences of signs. This work, however, includes the synthesis of the LSE classification phenomena, defining more complex elements than simple signs, such as *Classifier Predicates*, *Inflective* CCs and *Affixal classifiers*. The intelligibility of our synthetic messages was evaluated by LSE natives, who reported a recognition rate of 93% correct answers.
Key Words: Spanish Sign Language, Classifier Constructions, Automatic Synthesis, Deaf People
Category: I.3.7, I.2.7, H.5.2, K.4.2

1 Introduction

This work presents the first approach to representing and to synthesizing Spanish Sign Language (LSE) *Classifier Constructions* (CCs) using an LSE synthesizer [López Colino 2009, López-Colino and Colás In press].

Sign Language (SL) is not universal; each community of Deaf people has developed its own sign language. Many projects have focused on the translation and synthesis of a single SL, such as American SL (ASL) [VCom3D 2009], South African SL [van Zijl and Barker 2003], Greek SL [Karpouzis et al. 2007], LSE [López et al. 2006] and British SL [Marshall and Sáfár 2003]. The eSign project [Zwiterslood et al. 2004], as a European approach to SL synthesis, has focused on different European SLs, such as British SL, German SL and the Netherlands SL [Elliott et al. 2008, Kennaway et al. 2007]. There are also approaches that focus on multi-lingual solutions, but they are not based on phonetic descriptions of signs [Jemni and Elghoul 2008]. Although there are many references to SL synthesis, they focus on sign-based synthesis and omit the CCs.

The classification phenomena in LSE are quite complex. We present a simple example: consider the complexity of automatically generating the construction that a signer would use for this description: "put the suitcase in the luggage rack". The first element to be signed is the object of the action, SUITCASE; then the signer will present the position of the luggage rack with one hand using an 'L' configuration, because the luggage racks are rectangular; this hand will be placed in an upper position, because this is the position in which luggage racks are placed in trains and planes. Simultaneously, the other hand will use a fist-like configuration, simulating the grip of the suitcase's handle and move this second hand to the same height as the first one. Consider the world knowledge necessary for a machine translation (MT) module to generate this description from the original sentence; it is obviously a large problem to be addressed.

CCs are a relevant element in SL messages that should be considered in the process of translation and synthesis. It is not appropriate to omit CCs in synthetic messages. Although there are no studies on the appearance of these units in LSE, studies of ASL report that CCs occur about once per minute [Morford and MacFarlane 2003]. These structures are the preferred approach of Deaf people for expressing information about the spatial disposition of the elements mentioned in a discourse. This work seeks to define the way these elements are represented and synthesized.

Currently, all efforts at machine translation from Spanish to LSE obtain a sequence of signs to be synthesized [San Segundo et al. 2008a,b]. However, these projects do not generate the spatially and semantically complex CC structures included in LSE utterances. For example, the sentence "The book on the table" is represented in LSE using the following sentence: BOOK TABLE THAT CL "the book on the table" [see Fig. 1]. The construction CL "the book on the table" is a classifier predicate, with the passive hand representing the table (this hand is horizontal with the fingers extended) and the active hand representing the book. This construction is quite iconic, but it is not arbitrary. Section 2 will present the linguistic foundation of these constructions.

(a) BOOK (b) TABLE (c) THAT (d) CL "book over the table"

Figure 1: Signed sequence for the sentence "the book over the table".

Although the non-manual component (NMC) is present in most elements of signed discourse, classifier constructions show intensive use of the NMC. The NMC is composed of all the body and facial elements that, together with the manual components, create the signed communication [Herrero Blanco 2009]. These elements include: the eyebrows, the eyes, the eyelids, the tongue, the mouth (including the lips and the teeth), the head, the shoulders and the torso. All of these elements are present during the signing, altering different aspects of the morphology, syntax and semantics. The NMC is usually omitted from synthesized SL due to the complexity of its management. The approach to SL synthesis used in this work includes the synthesis of the NMC.

1.1 Context of this work

The work is presented in the context of the *MaTSyLSE* project (Machine Translation System for LSE). The *MaTSyLSE* project is a research project dealing with speech recognition, machine translation from Spanish to LSE and the synthesis of LSE messages [see Fig. 2]. This paper addresses the relevant elements of the notation created for describing LSE messages: the High Level Signing Markup Language (HLSML)[1] [López-Colino and Colás 2009], the modifications applied to the LSE synthesis module [López-Colino and Colás In press] and the evaluation of the resulting messages.

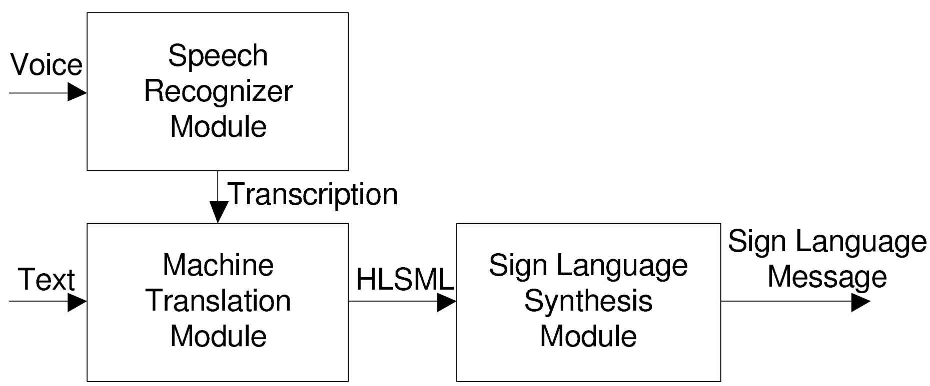


Figure 2: Diagram of the main modules of the *MaTSyLSE* project.

There are several examples of SL translation and synthesis modules in the literature. However, the number of web sites and applications that integrate

[1] The HLSML is an XML-based notation, its DTD document and tree structure can be found at `http://www.hctlab.com/research/hci/hlsml/`.

these accessible features is very low. The Spanish government enacted several laws [Gobierno de España 2003, 2007] that, ensure the same rights to LSE as to Spanish and require official web sites to provide signed content in order to make them accessible for Deaf people. Sign Language natives have a low reading comprehension level. Hence, Deaf people cannot fully understand text-based web pages. Today, after some years, these laws cannot be obeyed for technical and economic reasons. Deaf people reject synthetic signed messages. One reason for this rejection is that CCs, an important element in signed messages, are not included in current synthetic signed messages.

This work describes an approach to include these CCs in synthetic signed messages by means of a high level notation. This notation allows us to represent signed messages including signs, as in other approaches, and also including CCs, specifically those described for LSE. The syntheses of these CCs have been evaluated by LSE natives, who reported a recognition rate over 93%. A evaluation protocol has been developed for this work and the evaluation contents were validated by the FCNSE (Confederation of Spanish Deaf People), ensuring the reliability of the use of LSE during the tests[2].

1.2 Structure

This paper is structured as follows: [Section 2] reviews the current theory of CCs in LSE and defines the four kinds of CCs described for LSE. [Section 3] summarizes the state of the art in SL synthesis. This section presents the only existing approach to CC synthesis, designed for ASL, and discusses its limitations and the problems that prevent us from applying it to the synthesis of LSE CCs. [Section 4] presents our approach to the definition and the synthesis of CCs in LSE. [Section 5] describes the evaluation process [see Subsection 5.1] and obtained results [see Subsection 5.2]. Finally, [Section 6] summarizes this work and describes our future research objectives.

2 Linguistic Foundations

Since the initial research on SL phonology by Stokoe [Stokoe 1960], a variety of different phonetic models have been presented. The most extensive model, which has been applied in several different synthesis systems, is a parametric model [Rodríguez González 1992, Sandler and Lillo-Martin 2006]. This model defines seven different phonologic parameters: i) Configuration, or hand shape, ii) Orientation, the direction of the hand, iii) Location, or the position of the hand in the frontal plane, iv) Plane, the distance between the hand and the

[2] The same way the "Real Academia de la Lengua Española" or the "Académie française" watch over the correct use of Spanish or French, the FCNSE is the main LSE linguistic entity responsible for its study and normalization.

body, v) Contact Point, or active joint of the hand, vi) Movement of the hand and vii) a Non-Manual component (NMC), which includes facial expressions, head rotations and body movements. The HamNoSys notation [Prillwitz et al. 1989, Hanke 2004] was created to describe signs using this model.

Just as a variation of a manual parameter can modify a sign (see next section), the NMC can be also modified to alter the meaning of the sign: the temporal aspect of the verb can be defined by means of the value of the NMC. The signer can express agreement with the subject or the object of the verb by gazing towards the point where the referent has been placed inside the signing space. These realizations of the NMC are performed in a separate channel independent of the manual channel; the realizations can be simultaneous or sequential to the manual realization. The solution proposed in this work for the definition and synthesis of signed contents allows the management of both the manual and non-manual components of the signed communication.

A similar phonetic model was proposed by Herrero Blanco for LSE [Herrero Blanco 2009]. This second model merges the positions of the hands ("location" and "plane") in a single parameter and defines two kinds of movements: internal movements and external movements. The phonetic model considers variations of the "configuration" and the "orientation" as internal movements; a variation of the position of the hands is defined as an external movement. Both types of movements are defined using phonemes of the "movement" parameter. However, both models describe the locations of the hands using discrete anatomic references.

The first model has two main advantages. First, the independence of the parameters allows phonologic operations due to morphology (inflection, flexion, repetition ...) or syntax (the NMC is used to express questions, negations, etc.). Second, considering the "plane" and "location" parameters separately both reduces the number of phonemes and enables the morphologic modification of the "plane". This parameter is related to the temporal aspect, using different planes to refer to past, present or future actions. The synthesizer can easily merge the location and plane, thus reducing the number of units to be stored.

Sign Language messages include different kinds of elements: Spoken language influenced spelling sequences using the *fingerspelling* dictionary, *dictionary signs*, inflective constructions and flexions used for morphology, *Classifier Predicates*, etc. The purpose of this section is to present to the reader the complexity of the different elements in a signed message and to justify the proposed approach to the representation and the synthesis of CCs. Some authors rely on semantics or meaning to define and to categorize the classifiers. Herrero Blanco presented a classification of the CCs for LSE [Herrero Blanco 2004b]. This classification suggests four different kinds of CCs, based on the structure of the classifiers. The synthesis process relies on the structure (both phonological and syntactical), so we base our approach on it.

- The first elements, *fingerspelling* and *dictionary signs*, are the simplest ones, compared with the other more complex elements. Both *fingerspelling* and *dictionary signs* have a well-known phonetic description, which can be found in a dictionary. The *fingerspelling* is an alphabet representation by means of signs. Each letter is represented using a hand shape and an orientation. Most letters are static one-handed signs, but in some cases, the letters require a simple animation (wrist rotation or hand displacement). The *dictionary signs* (lemmas) represent concepts. The performance of these elements is more complex than *fingerspelling*. The descriptions of the base forms of the signs are stored in the lexicon. This lexicon is implemented as a relational database [López-Colino and Colás In press].

- The *Classifier Nouns* are *dictionary signs* that classify the following sign in the sentence. These signs can be used independently. with their own meanings, but they can be also used before another sign, to define a new concept by means of the classification phenomenon, e.g., the LSE sign WATER acts as a *Classifier Noun* when preceding different signs: WATER + PULLEY produces "well"; WATER + ROAD produces "river".

  This construction is morphologically similar to the construction of compound nouns (e.g., blackboard, bedroom; in English. WATER + EARTH is used for "mud", in LSE) and derivative nouns (e.g. strawberry, raspberry; in English. PERSON + SPAIN is used for "Spanish", in LSE). The difference between these two constructions and the *Classifier Noun* construction is purely semantic.

- The *inflective constructions* [Herrero Blanco 2009] imply some kind of phonologic operation on the base form of a sign. This operation implies a modification of the value of one or more phonologic parameters of the original sign. Such constructions can be found in different contexts, e.g., by modifying the orientation of the hand when signing a verb, the signer can modify the indirect object of the action. When I want to express "give something to you", my hand will point to the front: when I express "give something to me", my hand will point to my chest. By modifying the movement of the sign, the signer can represent the temporal aspect of the sign: already finished, a repetitive action, a continuing action, etc. The plural can be represented by modifying the configuration and using the corresponding numeral handshape (2-5). This is similar to the vowel change in English words to express plurals (e.g., man-men, woman-women, etc.).

  The *Inflective* CC is an inflective construction. The most commonly modified parameter is the configuration parameter of the sign, but the other parameters can also be modified by this type of construction. The objective of the construction is to add information to the original sign, usually related to the object of the action. The configuration can be modified using the classifier configuration[3] of the object of the sentence to define this verb-object relation (e.g., the sentence "I give you a book" in LSE will be represented by BOOK GIVE-cl:BOOK, meaning that the configuration of the sign DAR (give) is replaced by the configuration representing the book). A similar modification is used to express the locative: the initial position of the sign is modified to match the position of the sign that described the location of the action (e.g. the sentence "to climb a mountain" in LSE will be represented by MOUNTAIN TO_CLIMB-cl:MOUNTAIN). The initial position of the verb climb in this sentence is displaced to the location where the sign MOUNTAIN was signed.

In the synthesis of an inflective construction, the number of possible inflective modifications that a sign can receive is extensive. Hence, creating a full-form lexicon would require great efforts and storage capacity. We consider it a better approach to store the lemma of the sign, as we have done for the *dictionary signs*, and we propose a method to describe this kind of constructions and program the synthesizer to perform the required phonological operations.

– The *Classifier Predicates* (*CPs*) are complex elements that can be found in a SL message. These constructions are very similar to iconic realizations, but several authors describe a linguistic basis for them. There are many published works describing this kind of element in a signed message [Schembri 2003, Schembri et al. 2005, Liddell and Metzger 1998, Liddell 2003, Cogill-Koez 2000]. These constructions depict the spatial distribution, size, orientation and relation between different referents introduced during the signed conversation. Whereas the *Inflective* CCs are modifications to the phonetic description of a sign, the *CP* are productive elements, allowing great freedom to the signer in their execution.

  These units are found in the literature as *CP* [Liddell 2003] or *verbs of motion and location* [Supalla 2003]. Several authors consider these *CP* as linguistic units that are equivalent to other CCs. However, Emmorey and Herzig showed that although the use of the hand shapes follows a morphemic approach, hand positioning does not share this approach, thus showing that "the depictive use of a signing space is a gestural component of classifier constructions, rather than a morphemic component." [Emmorey and Herzig 2003]. Different authors argue about the linguistic nature of this CC. As we are not interested in this discussion, we will focus on the morphology of

[3] The classifier configuration is a hand-shape that represents the category of the object. It is not the hand-shape of the sign that represents the object.

these constructions and the differences between them and other structures found in a signed message. The most important difference is related to hand positioning and movement: hand positioning cannot be described by the discrete set of positions defined by the location and plane parameters, as it requires an analogical positioning approach. Something similar applies to hand movements; the depiction of a real scene may imply any arbitrary movement, which has to be described. The other parameters are also relevant in this construction. However, their management does not differ from that used in *dictionary sign* synthesis.

When approaching the synthesis of a *CP*, it would obviously be very difficult to predefine all possible *CP* and store them in a corpus. These elements depend on the message contents, so when processing the message-to-be translated, the description of a *CP* must be created. This is a big problem in MT research, with few published works [Huenerfauth 2006]. We will not discuss this issue in this paper. Our objective is to provide an interface that will allow future MT systems to describe *CPs*, so they can be synthesized using our SL synthesis module and so SL natives can evaluate the results.

– The last kind of CC units proposed by Herrero Blanco are the *Affixal classifiers*. One of the differences between SLs and Spoken Languages is the productive organs: speech is generated using the vocal tract. Humans only have one vocal tract, so it is impossible to produce parallel utterances. Sign Languages use the whole body to produce the signed message. The hands and the NMC are three different channels that can generate contents simultaneously, and they can be combined to express different concepts at the same time. The *Affixal classifier* merges the *Inflective* CC and the *CP*; while the passive hand is performing a *CP*, the active hand simultaneously performs a *dictionary sign* whith a modified orientation, location and plane. This modification results in an alteration of the performance of the sign, either by modifying the last syllable of the *dictionary sign* in a disyllabic sign or by including a new syllable for the monosyllabic signs[4]. In both situations, the new constructions have the active hand oriented and positioned in relation to the passive hand. A similar construction in an oral language is the Spanish translation of "give it to him", which is the single Spanish word "dárselo", concatenating the verb "dar", the pronoun for the indirect object "se" and the pronoun for the direct object "lo".

  These kind sof CCs can be observed in the construction "to look at an orange", which can be described using the following sequence of actions (using a right-handed avatar): i) the sign ORANGE (as a fruit) is performed using its base form. ii) The left hand establishes the spatial position of the orange,

[4] The concept of syllable applied to SL was presented in [Brentari 1996].

described as a *CP*. iii) The right hand performs the sign TO_LOOK_AT, but the final position of the right hand is modified to be located near the left hand, which is the virtual position of the orange; this modification can be described as an *Inflective* CC of several parameters of the base form of TO_LOOK_AT.

The *Affixal classifiers* require us to describe behaviors of the hands. It must be noted that the behaviors of the hands are not independent, as the modifications made to the *dictionary sign* are described by means of the passive hand, which performs a *CP*.

- The *pointing signs* are pointing gestures (the hand points to a position in the signing space). These signs serve a number of functions [Sandler and Lillo-Martin 2006], such as determiners and pronouns; they are also related to nominals to define the location to be associated with them. This location is used by subsequent signs (starting the sign or orienting the sign to that position) to express related information. The different linguistic functions are relevant to the MT process. However, during the synthesis process, all of them are assimilated to a pointing gesture.

- Signers use different techniques to represent a dialog or interaction between two or more people, i.e., *role shifting.* Lillo-Martin [Lillo-Martin 1995] considers that the signer modifies elements of the NMC to represent the different roles: shoulder and head movements, gaze direction and facial expression. However, Liddell & Metzger [Liddell and Metzger 1998] did not find consistency in these modifications, so they should not be considered as the only way a signer represents role shifting. Engberg-Pedersen [Engberg-Pedersen 1995] observed that along with the changes in the face expression, there is also a change of references in the signing space when placing objects (e.g. the object placed to the left of character 1 would be placed to the right of character 2, if these two characters are face to face).

3 State of the Art

There are several examples of SL synthesizers in the literature. The technology used for SL synthesis comes from different approaches: the ones based on video sequences [Solina et al. 2001], the ones based on animations captured from a human signer [Sagawa and Takeuchi 2002, Loomis et al. 1983, Bangham et al. 2000] and the ones based on phonetic descriptions of the signs used for generating an avatar animation. The previous section has presented the morphologic modifications that the CCs impose on the *dictionary signs*, so the only valid approach to the synthesis of the CCs is the one based on phonetic descriptions of the signs. This approach is also known as “parametric synthesis”.

Most SL synthesizers use standard symbolic notations to describe signs. Notations such as HamNoSys [Prillwitz et al. 1989, Hanke 2004] and SignWriting [Sutton 1974] are graphic representations of SL and have computer-friendly versions: SiGML [Kennaway et al. 2002] for the HamNoSys notation and SWML [Rocha and Pereira 2004] for SignWriting. Gesture synthesis for these projects is a direct conversion from SWML or SiGML into VRML. Grieve-Smith [Grieve-Smith 2002], uses the Stokoe notation [Stokoe 1960] to define the signs. There is also another representation system, called "Szczepankowski's gestographic notation" [Francik and Fabian 2002] used in Polish Sign Language. This is a textual notation as it uses regular ASCII characters, so it is computer friendly. However, it does not represent all of the sign parameters. This problem also occurs in SEA notation [Herrero Blanco 2004a], which has been developed with a focus on LSE. All of these notations require considerable knowledge of the SL structure and learning to be used.

When focusing on CC synthesis, the number of references that can be found is reduced to one work [Huenerfauth 2004, 2006], which is the first and only approach to ASL *Classifier Predicate* automatic translation and synthesis. This approach used a software system to obtain a 3D scene, described by natural language (English): AnimNL [Bindiganavale et al. 2000, Schuler 2003, Badler et al. 2000]. The information obtained from the AnimNL program was used as basis for the automatic synthesis of the ASL *CPs*. This program uses a library of Parameterized Action Representations (PARs) as templates to describe the scene and the actions of all the elements that populate that scene. The AnimNL was created to process scene descriptions found written in English. There is not an equivalent software system for Spanish that could be used as a basis. Huenerfauth's work only provides a partial solution for *CP* (e.g., the previous suitcase example cannot be described). However, we must remark another important contribution referring to SL synthesis, which is a description of parallel actions of the NMC while performing *dictionary signs* [Huenerfauth 2005]. Although that contribution cannot be applied to *Affixal classifiers*, it describes a similar parallel signing behavior.

The first approach to Spanish to LSE machine translation [San Segundo et al. 2008a,b, Baldassarri and Royo Santas 2007] was based on a *dictionary sign* translation approach, omitting the translation and processing of CCs.

4 Representation and Synthesis of LSE *Classifier Constructions*

This section describes how the elements of a signed message, presented in [Section 2], are defined in the input notation. It also describes process of synthesizing *dictionary signs* [López-Colino and Colás In press] can be altered to include the classifier constructions in the synthetic message.

4.1 Input notation

The High Level Signing Markup Language (HLSML) notation that we have created was designed with the following four objectives: i) Generate an XML-based notation that could be used by both people with some LSE knowledge and by a MT module. ii) Describe LSE messages, including all their possible elements (*fingerspelling* dictionary, *dictionary signs*, CCs, etc.) in the same specification. iii) Allow the use of modifiers to change the representation of a sign (inflective modifications). iv) Describe the parallel behavior of NMC and the *Affixal classifiers* in the signing process.

The representation of the different elements in a signed message, using the HLSML specification requires two different approaches, one for the *dictionary signs* (element <sign>) and *fingerspelling* (element <spell>) and another for the CCs and the morphologic modifications, i.e., one approach for the elements that have a static definition (*dictionary signs* and *fingerspelling*), which is stored in the database and another for those elements that are described depending on the message. HLSML only establishes a reference to the sign glosses or the letters in the *fingerspelling* alphabet, as HLSML defines a *fingerspelling* sequence by means of the word to be spelled. The other approach, used for the message dependent elements, require the phonetic description to be included in the input message. In this case, the input notation refers to the phonemes stored in the relational database.

4.2 Description of the Non-Manual Component

We have already stated the importance of the NMC for the signed communication. The system handles two different approaches to the description of the NMC: the first one is to store the related information in the database along with the other manual-related parameters. This approach is relevant for those signs that require a NMC realization in their base form. The second one is to describe the NMC channel in the input notation. This second approach is used to define NMC that affect to the sentence or NMC descriptions related to the message-dependent elements (e.g., classifiers) that cannot be described in advance.

As we have stated before, the NMC is divided into many sub-channels (head, shoulders, different facial parts, etc.). HLSML describes a sequence of NMC animations using the tag <nonManualSequence>. This tag defines the duration of the NMC sequence and includes several <phoneme/> tags, each of which is used to state the value of the NMC sub-channels along the animation. [Fig. 3] shows a simple example stating that the avatar must blink its eyes.

Another feature of the HLSML notation that affects the NMC is the visual speech capabilities of the synthesizer. The synthesizer was constructed as a common solution for deaf people, as a SL synthesizer, and for hard of hearing people,

```
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
...
 <nonManualSequence time="500">
  <phoneme value="eyes_open"
    fraction_ini="0" fraction_end="30" />
  <phoneme value="eyes_closed"
    fraction_ini="40" fraction_end="60" />
  <phoneme value="eyes_open"
    fraction_ini="70" fraction_end="100" />
 </ nonManualSequence>
...
```

Figure 3: Description of a blink. The whole sequence's duration is $500ms$, as stated by the value of the "time" argument. The animation is composed by three steps; the duration of each step is described using percentages of the sequence's total duration. Hence, the first <phoneme> last from the $0ms$ instant (fraction 0) to the $150ms$ instant. The second step ($200ms$ to $300ms$) states that the avatar's eyes must be closed. The animation segment between $150ms$ and $200ms$ (30% to 40%) is the transition between the open eyes and the closed eyes. Finally, the last step states that the avatar must open the eyes again.

using visual speech as a complement for visual interfaces. These two features are independent but the synthesizer allows both of them. To define a visual speech animation, the HLSML notation includes the <talk> element, which defines the speech segment, and the <viseme> that is used for describing the sequence of visemes in the speech.

Finally, we also stated that the avatars gaze is important for correct signing. The HLSML notation includes two XML elements to define a point in the signing space at which the avatar must gaze: the <headLookAt> and <eyesLookAt> elements. These elements may use either a coordinate in the signing space or a point of the anatomy, similar to the location parameter, (e.g., the avatar has to gaze to its right little finger).

These elements will be present during the description of several LSE CCs, as we will see now.

4.3 Representation and Synthesis of *Classifier Nouns*

This CC is composed of a sequence of two *dictionary signs*: the first sign adds its meaning to the second one in order to create a new concept. The synthesis of a *Classifier Noun* does not modify the synthesis process of a *dictionary sign*, because it is equivalent to a sequence of two independent *dictionary signs* retrieved from the relational database.

The notation HLSML includes the <sign> element to define the synthesis of a *dictionary sign*. The value of the "value" attribute defines the gloss of the sign to be synthesized. A *Classifier Noun* is defined using two consecutive <sign>

elements and, if required, the time between these two *dictionary signs* can be modified using the <timeInter> element. [Fig. 4] shows the example proposed in [section 2] for a *Classifier Noun*.

```
<!DOCTYPE hlsml SYSTEM
 "hlsml.dtd">
<hlsml>
 <sentence language='lse
   tag ='standard'>
  <sign value="AGUA" />
  <sign value="POLEA"/>
 </sentence>
</hlsml>
```

(a) HLSML definition. (b) Sign AGUA (water) (c) Sign POLEA (pulley)

Figure 4: Example of a *Classifier Noun*, where the sign AGUA (water) and the sign POLEA (pulley) merge their meanings to create the concept "pozo" (well).

4.4 Representation and Synthesis of *Inflective Classifier Construction*s

The *Inflective* CC is a phonological modification applied to *dictionary sign*s stored in the lexicon. Hence, the HLSML notation describes these inflective constructions as modifiers of the <sign> using the <inflectiveModification>. These constructions require us to define both the modified parameter and its new value. The modified parameter is defined as an attribute of <inflectiveModification>. The new phoneme can be defined either by stating the phoneme that must be used (<phoneme/>) or by stating a sign whose phonetic description is involved with this parameter. The synthesizer's database (see [López et al. 2006, López-Colino and Colás In press]) includes the phonetic definitions of the *dictionary signs*, but it is also used for storing other information required for LSE synthesis. The database contains templates used for LSE synthesis, such as the classifier configurations and orientations related to a gloss (e.g., the classifier of a person uses the extended pointing finger hand shape and the orientation states that the finger must point to the ceiling; a car uses a hand shape that consists of all fingers extended, and the hand must be horizontal). These templates do not include the definition of every parameter, so they cannot be synthesized in isolation; rather, they are used as auxiliary elements.

[Fig. 5] depicts the "to climb a mountain" example proposed in [section 2]. The definition of the location parameter of TO CLIMB is modified using the location phoneme of the sign MOUNTAIN. The database stores the base forms of both signs. During the synthesis process, this module retrieves the required description for each parameter.

```
1 <!DOCTYPE hlsml SYSTEM "hlsml.dtd">
2 ...
3  <sentence>
4   <sign value="mountain" />
5   <sign value="to_climb"
6    <inflectiveModification
7      value="location">
8     <sign value="mountain" />
9    </ inflectiveModification>
10  </ sign>
11 </ sentence>
12 ...
```

(a) HLSML message

(b) Base form TO CLIMB

(c) MOUNTAIN, line 4

(d) TO CLIMB cl:mountain, line 5 to 10

Figure 5: Example of an *Inflective* CC, where the sign SUBIR (to climb) is modified by the sign MONTAÑA (mountain), so the mountain is the locative argument of the verb.

4.5 Representation and Synthesis of *Classifier Predicates*

Classifier Predicates are spatially-related descriptions included in the message. These elements are used to define the spatial positions, spatial relations and movements of the elements cited in the conversation. The *CP* can be also used to describe relations between different elements not related to spatial properties (e.g., the subject and the object of a sentence can be related using *CP*). As their description is based on semantic information obtained from natural language processing, there is no possibility of storing the final representations of *CPs* final representation, so these constructions must be defined and generated dynamically.

The most relevant difference between *CP* and the *dictionary signs* for the synthesis is the positioning of the hands. Though the *dictionary signs* use location and plane phonemes related to discrete anatomic references, the *CP* does not use the anatomic reference approach to hand positioning. The *Classifier Predicate*s require a different approach because of their productive nature; this approach must also be related to the representation of real scenes described using Cartesian coordinates, not the body reference approach. Finally, the adopted approach for the *CPs* should be able to handle different degrees of accuracy. There will be representations that will only require placing three distant objects, whereas others will require placing ten. The approach we have developed for the synthesizer is as follows: the signing space uses the position of the chest ($\overrightarrow{Pos_{chest}}$) as the origin of the coordinate system (the point with coordinates $(0, 0, 0)$). The synthesizer may use different avatars with different scales, so the coordinate approach must be relative to a predefined dimension of the avatar. We have chosen the length of the upper arm bone as the length unit (Arm_{length}). Each classifier predicate will require a different accuracy level for the hand positioning. This accuracy is achieved by dividing the length unit as required ($Arm_{division}$). The final 3D position ($\overrightarrow{Pos_{final}}$) in the virtual scene for a given coordinate <coordinate horizontal="i" vertical="j" frontal="k"/> is obtained using (1). This approach allows us to define any position in the articulatory space, independently of the avatar size [Fig. 6].

$$\overrightarrow{Pos_{final}} = \overrightarrow{Pos_{chest}} + \frac{Arm_{length}}{Arm_{division}} (i, j, k) \tag{1}$$

The <classifierPredicate> element is used to describe a *CP* in HLSML. This XML element only defines the value of "armDivision". As mentioned above, the most common actions are placing objects and moving objects; the <placeObject> and the <moveObject/> are used for these tasks. Both include the semantic category of the referent (obtained during the MT process), the duration of the construction, the hand that performs the *CP* and either the coordinates where the object is placed or a sequence of coordinates describing the movement's trajectory. The semantic category is represented by a hand-shape, with a default orientation and contact point. These values are stored as a template in the relational database. The synthesizer retrieves this template to generate the message.

If a more detailed description is required, the <staticPosition> describes a static position of its descendant elements. The <staticPosition> allows one <hand> element to be included for each hand. The <hand> defines the value for

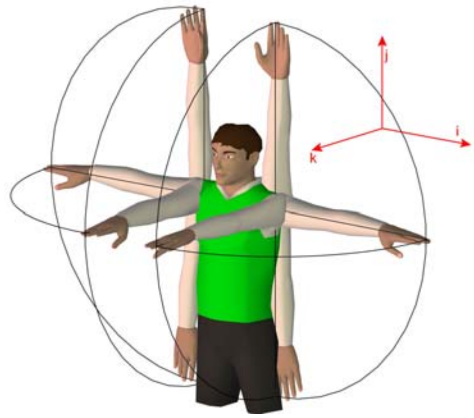


Figure 6: Every possible position within the articulatory space can be defined using the proposed coordinate system.

the configuration, orientation and contact point parameters and a <coordinate>. The duration of the transition between two consecutive <staticPosition> is defined using the <positionTransition> element.

Consider that the combination of <staticPosition> + <hand> is equivalent to a <placeObject>. Analogously, a sequence of <staticPosition> and <positionTransition> is equivalent to a <moveObject>. However, the latter include a semantic category, which allows the synthesizer to retrieve the values for the configuration, orientation and contact point parameters from the database.

The *CP* construction also requires a definition of NMC. The <classifierPredicate> allows the inclusion of <nonManualSequence> elements, and the <staticPosition> can include both <headLookAt> and <eyesLookAt> elements. Hence, we can describe the NMC elements that must be represented in a *CP*. An example of this construction is presented in [Fig. 7].

4.5.1 Pointing Signs

Although they are used for different lexical functions, the pointing signs are pointing gestures. A pointing sign can be described using HLSML and the structure <staticPosition>: defining the configuration as the pointing hand-shape, the contact point is the end of the pointing finger and the <coordinate> defines the position that the avatar will point to.

4.5.2 Role shifting

There are two main elements in the role shifting process: variations of the NMC and changes in the reference point when positioning referents in the signing

```
1 <!DOCTYPE hlsml SYSTEM "hlsml.dtd">
2 ...
3 <classifierPredicate armDivision="4">
4  <placeObject value="person" time="200"
5    side="right">
6   <coordinate vertical="0" horizontal="-4"
7    frontal="3" />
8  </placeObject>
9  <positionTransition time="400" />
10 <placeObject value="person" time="200"
11   side="right">
12  <coordinate vertical="0" horizontal="1"
13   frontal="3" />
14 </placeObject>
15 </classifierPredicate>
16 ...
```

(a) HLSML message

(b) Initial Position, line 4

(c) Transition ($400ms$), line 9

(d) Final position, line 10

Figure 7: Example of a CP, where the avatar describes a person walking along a linear path, from its right to the left.

space. We have presented the way to define variations to the NMC using the <nonManualSequence>. The position in the signing space is described in the input message. Although the synthesizer will correctly represent role shifting, it has to be defined during the translation process.

4.6 Representation and Synthesis of *Affixal Classifier Constructions*

The *Affixal classifiers* merge the definition of the *Inflective* CCs and the *CPs*, so their description requires the definition of parallel behavior of the different CCs that are required for the definition of these *Affixal classifiers*. The element <compound> is used to define parallel behavior of its child elements.

The <affixalClassifier> defines this kind of modification for a <sign>. The difference between the <inflectiveModification> and the <affixalClassifier> is that the latter does not describe the new phonemes that must be used; it only defines the part of the non-dominant hand that must be used as a reference. An example of this construction is presented in [Fig. 8].

5 Evaluating LSE *Classifier Constructions*

The common approach to evaluating the synthetic signed messages is a user-based evaluation. This approach is mandatory due to the semantic interpretation that is required to understand the CCs, which makes an automatic evaluation approach impossible. The most demanding users are natives of the evaluated language because they perceive more details in the message than SL interpreters or non-native SL users. The evaluations were designed to measure the understanding of the synthetic messages.

The evaluations must simulate the real environment of the system. We have stated before that many official web pages should be accessible using LSE; for this reason, we have designed an evaluation approach to emulate access to web contents. In this scenario, the user does not have the support of a SL interpreter, the evaluator is not in the room, and the user's doubts cannot be resolved by another person. The only information available to the user is the synthetic signed contents and the web site contents (mostly text). The approach presented by other researchers [Huenerfauth et al. 2008] relies on a SL interpreter to instruct the user about the experiment. This communication between the user and the interpreter does not exist in normal access to web contents. For this reason, our evaluation protocol only relies on a brief and textual description of the experiment to inform the user of what to do. Furthermore, the evaluations focus on the understanding of the synthetic signed messages, so the test must avoid including any translation process, including the manifestation of the answer by the user. For this reason, when we evaluate the constructions that depict a scene,

```
1 <!DOCTYPE hlsml SYSTEM "hlsml.dtd">
2 ...
3 <sentence language="lse">
4  <sign value="orange"/>
5  <classifierPredicate armDivision="3">
6   <placeObject value="spherical" time="200"
7     side="passive">
8    <coordinate vertical="1" horizontal="2"
9     frontal="3" />
10   </placeObject>
11  </classifierPredicate>
12  <compound>
13   <sign value="to_look_at">
14    <affixlaClassifier>
15     <element value="passive_hand"/>
16    </affixlaClassifier>
17   </sign>
18   <classifierPredicate armDivision="3">
19    <placeObject value="spherical"
20      side="left">
21     <coordinate vertical="1" horizontal="2"
22      frontal="3" />
23     </placeObject>
24   </classifierPredicate>
25  </compound>
26 </ sentence>
27 ...
```

(a) Base form of TO_LOOK_AT

(b) HLSML message.

(c) Sign ORANGE, line 4

(d) Object positioning, line 6

(e) Modified sign TO_LOOK_AT, line 13

Figure 8: Example of an *Affixal classifier*, in the sentence "to look at an orange". Figures 8(c), 8(d) and 8(e) show the sequence of the sentence. where the sign MIRAR (to_look_at) is oriented towards the position of the left hand, which represents the object of the "look at" action in the articulatory space.

we do not rely on text descriptions of the scene; the evaluation approach presents different images to the user and asks him/her to choose the scene described by the synthetic signed message.

The morphologic structure of the LSE CCs and their semantic representation require the use of two different experiment strategies: one for the *Classifier Nouns*, similar to the *dictionary signs*, and another for the more complex *In-*

flective CC, the *Classifier Predicate* and the *Affixal classifier*.

5.1 Experimental Setup

The aim of the evaluations is to measure understanding of the signed messages in a real web usage environment. For this reason, every evaluation was created as a web form in which each question was presented in a different web page to the user. The first page of the evaluation used a simple form to ask the user some information about his/her age, gender and to confirm that s/he is a LSE native. After gathering this information, the system briefly described the experiment and gave the user instructions. We presented the users with two different evaluations, the first one focused on the *Classifier Nouns* and the second one on the other LSE CCs:

The first evaluation focused on the *Classifier Noun* construction, a sequence of two *dictionary signs*. Hence, the evaluation of this kind of CCs has been designed as an isolated sign recognition experiment. The experiment was divided into two different sets of questions, each containing 20 different signs, which are not repeated between the two sets. The 40 signs list was composed of nouns, adjectives and verbs, comprising single and double handed signs. Each question contained a video generated with the LSE synthesizer; each video presented the avatar starting in a neutral gesture, performed an isolated sign, and returned to the initial gesture. In the first part of the experiment, corresponding to the first set of signs, the users were asked to write the sign presented by the avatar, if recognized, as the system allowed them to leave the answer blank. The second part of the experiment displayed the same kind of videos, but this time the questionnaire was a multiple-choice test; it presented the correct answer and four false alternatives. This time, the user was required to provide an answer in order to proceed to the next question.

The second evaluation focused on the more semantically-complex LSE's CCs. For this reason, instead of proposing a multiple-choice based on text descriptions of the scene for each possible answer, we used images to depict the different answers to each question. We have based our design on the questionnaires created by Huenerfauth. He proposed 10 scenes described using an avatar and proposed 3 different answers. Our evaluation consisted of 16 different questions, which included the last 3 kinds of CCs. Each question proposed 5 possible answers to the users. [Fig. 9] shows a question related to a *CP*, which describes a person walking from the church (placed on the left) to the house (placed on the right). Consider the different possibilities for this experiment: the kind of building (chosen between a house, a church and a tower), their position (left or right) and the movement of the person (left to right, right to left, staying between them, moving between them, etc.). Using images for this kind of test was the solution preferred by the FCNSE experts. They considered this approach better than using Spanish text descriptions, making the evaluations easier for the users. The 16 signed animations were validated by a LSE linguist from the FCNSE. This expert did not participate in the evaluation process.

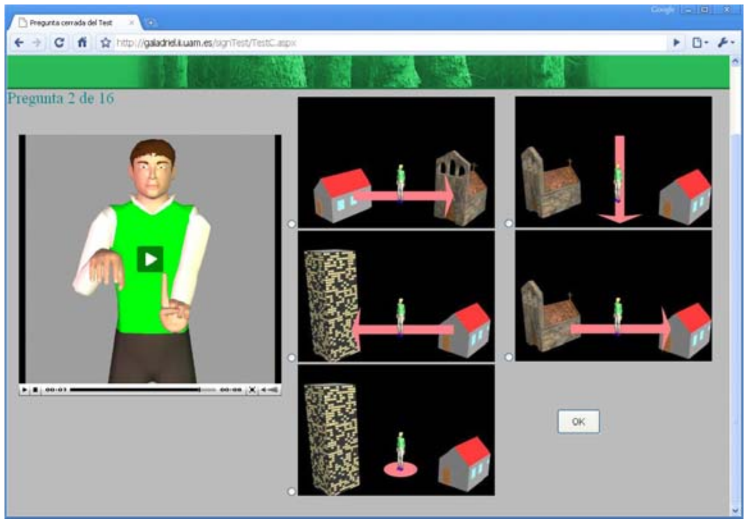


Figure 9: Screenshot of one of the questions used for evaluating the CCs intelligibility.

The evaluation group was composed of 11 LSE natives (7 males and 4 females), aged between 24 and 50 years old; two of these users work as linguistic experts on the FCNSE. It is important to note that all these users live in the same city, so the regional dialect variations that exist in LSE will not influence the results of the evaluations. The regional variations decrease the recognition rate, as the same concept is represented by a different *dictionary sign* depending on the dialect, as reported in [San Segundo et al. 2008a]. Although the users were LSE natives, they had medium-level knowledge of written Spanish, but enough to understand the written instructions of the evaluations. They are also experienced pc users, so we could send them the URLs of the evaluations by e-mail. Eight of the users were high school graduates, and the other three had a bachelor's degree.

This group does not fully represent the signing deaf community. However, we are only evaluating the performance of the signing avatar, when signing classifier constructions. Our aim is to evaluate whether a classifier described using HLSML and synthesized using our system, can be correctly recognized by signing people. The automatic generation of these constructions and the correctness of their usage in the machine translated messages will need to be evaluated with a larger group. This new group will represent the entire Deaf community. For these evaluations we will replace the initial text-based description using a signed video (recorded by a human interpreter).

5.2 Results

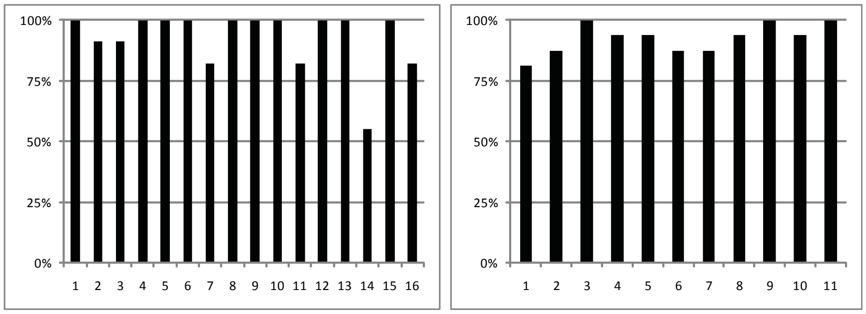
The results obtained for the *Classifier Nouns* evaluation, using the isolated-sign-like evaluations, reported a recognition rate of 96% of correct answers for the first set of signs and a rate of 98% of correct answers for the second test. In the first set of signs, we have considered the fact that the same sign can be used to represent different Spanish concepts (e.g. "water" and "to drink" are represented by the same sign). The isolated sign evaluation does not provide context information that could allow the user to identify the concept being represented. For this reason, we have considered as correct answers all the concepts that correspond to the represented sign.

The increase of the recognition rate from the first to the second test is due to two reasons: i) the testing approach of the second set of signs, the multiple-choice test, always includes the correct answer and does not allow the user to leave the question unanswered. This approach simplifies the answering as the user chooses the most similar answer. ii) The first set of signs is also the first time the synthetic contents are presented to the users. When the users start the second test, they have a brief experience with the avatar's appearance and signing style as the result of the training effect that the first set of signs presented to the users.

The results obtained in the second experiment (used for the evaluation of the *Inflective* CC, *CP* and *Affixal classifier*) present a recognition rate of 93% of correct answers [see Fig. 10]. Every scene described by the avatar by means of a CC reports a recognition rate of over 80%, except for Question 14. The recognition rate for Question 14 was 55% of correct answers. This last result contrasts with the results obtained for the other questions. This question corresponds to a *CP* depicting three cars parked side by side. Although the avatar displayed the spatial position of the cars correctly, we realized that the avatar described the scene from a specific point of view, placing the observer behind the cars. The correct answer for this question presented the three cars being observed from the side of one of them. This is the only question that presented the correct answer from a different point of view than the one used in the signed message, which justifies the result obtained.

5.3 Comparison with other CCs synthesis approaches

Finally, we compare our results with the results reported by Huenerfauth et al. in [Huenerfauth et al. 2008]. The experimental setup presented in our work



(a) Results obtained in each question. (b) Results obtained per user.

Figure 10: Graph of the obtained recognition rate for the evaluation of different *CPs*, *Affixal classifiers* and *Inflective* CCs.

is slightly different from the one reported by Huenerfauth: they proposed ten different scenes, both described by a *CP* and a signed English sequence (we used sixteen different scenes, all corresponding to CCs); their test approach is different, in that they show an avatar animation and three different videos representing different situations, of which only one is correct (we present one avatar animation and five different alternative scenes, used as possible answers); their experiment was performed by fifteen participants whereas ours was performed by eleven participants, all LSE natives. Their matching results report 85% of correct answers, and although we are aware that their experimental conditions are slightly different, our work reports an 8% improvement in the recognition rate.

6 Conclusion

We have presented a novel and functional approach to the description and synthesis of CCs in LSE, integrated into a LSE synthesizer [López-Colino and Colás In press]. These constructions are present in signed communication. Including them in synthetic signed messages will promote the acceptance of this kind of messages by deaf people. Therefore, using our LSE automatic synthesizer, we will be able to create signed messages with the same contents as the human ones. This work also presented how using HLSML, the notation described in this work, the representation of the CCs can be done.

We have presented an evaluation approach for measuring the recognition rate of the different LSE CCs. Using LSE native users, we have obtained an average recognition rate of over 93%, depending on the kind of the CC evaluated. Although it was not one of our evaluation objectives, we also observed the importance of the point of view when describing CCs. This is a relevant issue for the

CPs as they can be used to represent reality. The definition of the point of view should be considered in automatic CC generation, a project that is currently under development in the *MaTSyLSE* project.

Acknowledgments

Authors would like to acknowledge the FPU-UAM program for its support and the people from the FCNSE linguistic department who performed the evaluations.

References

Badler, N. I., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., joo Lee, S., Shin, H., Palmer, M.: "Parameterized action representation and natural language instructions for dynamic behavior modification of embodied agents"; AAAI Spring Symposium; 2000.

Baldassarri, S., Royo Santas, F. J.: "Traductor de español a lse basado en reglas gramaticales y morfológicas"; Proceedings of INTERACION; AIPO, 2007.

Bangham, A., Cox, S., Elliot, R., Glauert, J., Marshall, I.: "Virtual signing: Capture, animation, storage and transmission - an overview of the visicast project"; IEE Seminary on Speech and Language Processing for Disabled and Elderly People; 2000.

Bindiganavale, R., Schuler, W., Allbeck, J. M., Badler, N. I., Joshi, A. K., Palmer, M.: "Dynamically altering agent behaviors using natural language instructions"; Proceedings of the 4th international conference on Autonomous agents; 293–300; ACM, New York, NY, USA, 2000.

Brentari, D.: "Trilled movement: Phonetic realization and formal representation"; Lingua; 98 (1996), 1–3, 43–71.

Cogill-Koez, D.: "A model of signed language 'classifier predicates' as templated visual representation"; Sign Language and Linguistics; 3 (2000), 2, 209–236.

Elliott, R., Glauert, J., Kennaway, R., Marshal, I., Sáfár, E.: "Linguistic modelling and language-processing technologies for avatar-based sign language presentation"; UAIS; 6 (2008), 4, 375–391.

Emmorey, K., Herzig, M.: Perspectives on Classifier Constructions in Sign Languages; chapter 10: Categorical Versus Gradient Properties of Classifier Constructions in ASL, 221–246; Psychology Press, 2003.

Engberg-Pedersen, E.: Language, Gesture, and Space; chapter 7: Point of view expressed through shifters, 133–154; Lawrence Erlbaum Associates, 1995.

Francik, J., Fabian, P.: "Animating sign language in the real time"; Proceedings of 20th IASTED International Multi-Conference Applied Informatics; 276–281; 2002.

Gobierno de España: “Ley 51/2003, de 2 de diciembre, de igualdad de oportunidades, no discriminacin y accesibilidad universal de las personas con discapacidad”; B.O.E. 3 December (2003); (Spanish Law).

Gobierno de España: “Ley 27/2007, de 23 de octubre, por la que se reconocen las lenguas de signos espaolas y se regulan los medios de apoyo a la comunicacin oral de las personas sordas, con discapacidad auditiva y sordociegas.”; B.O.E. 24 October (2007); (Spanish Law).

Grieve-Smith, A. B.: “Signsynth: A sign language synthesis application using web3d and perl”; Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction; 134–145; Springer-Verlag, London, UK, 2002.

Hanke, T.: “Hamnosys - representing sign language data in language resources and language processing contexts”; S. B. . Heidelberg, ed., Proceedings of LREC; 1–6; Lisbon, 2004.

Herrero Blanco, Ángel.: “A practical writing of sign languages”; Proceedings of LREC; 37–42; Lisbon, 2004a.

Herrero Blanco, Ángel.: “Una aproximación morfológica a las construcciones clasificatorias en la lengua de signos española”; Estudios de Lingüística de la universidad de Alicante; 18 (2004b), 151–167.

Herrero Blanco, Ángel.: Gramática Didáctica de la Lengua de Signos Española (LSE); Ediciones SM, 2009; 1 edition.

Huenerfauth, M.: “Spatial representation of classifier predicates for machine translation into american sign language”; Proceedings of LREC; 24–31; Lisbon, 2004.

Huenerfauth, M.: “Representing coordination and non-coordination in an american sign language animation”; Proceedings of ACM SIGACCESS; 44–51; ACM, New York, NY, USA, 2005.

Huenerfauth, M.: Generating american sign language classifier predicates for english-to-asl machine translation; Ph.D. thesis; Computer and Information Science, University of Pennsylvania; Philadelphia, PA, USA (2006); adviser Mitch Marcus and Martha Palmer.

Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: “Evaluation of american sign language generation by native asl signers”; ACM Transactions on Accessible Computing; 1 (2008), 1, 1–27.

Jemni, M., Elghoul, O.: “A system to make signs using collaborative approach”; K. Miesenberger, J. Klaus, W. Zagler, A. Karshmer, eds., Computers Helping People with Special Needs; volume 5105 of Lecture Notes in Computer Science; 670–677; Springer Berlin / Heidelberg, 2008.

Karpouzis, K., Caridakis, G., Fotinea, S.-E., Efthimiou, E.: “Educational resources and implementation of a greek sign language synthesis architecture”; Computers & Education; 49 (2007), 1, 54 – 74.

Kennaway, R., Elliot, R., Glauert, J., Parsons, K. J.: "Sigml document type definition (dtd)"; Visicast deliverable; ViSiCAST Project (2002).

Kennaway, R., Glauert, J., Zwitserlood, I.: "Providing signed content on the internet by synthesized animation"; ACM Transactions on Computer-Human Interaction; 14 (2007), 15, 1–29.

Liddell, S. K.: Perspectives on Classifier Constructions in Sign Languages; chapter Sources of Meaning in ASL Classifier Predicates, 199–220; Psychology Press, 2003.

Liddell, S. K., Metzger, M.: "Gesture in sign language discourse"; Journal of Pragmatics; 30 (1998), 6, 657–697.

Lillo-Martin, D.: Language, Gesture, and Space; chapter 8: The Point of View Predicate in American Sign Languages, 155–170; Lawrence Erlbaum Associates, 1995.

Loomis, J., Poizner, H., Bellugi, U., Blakemore, A., Hollerbach, J.: "Computer graphic modeling of american sign language"; ACM SIGGRAPH Computer Graphics; 17 (1983), 3, 105–114.

López, F., Tejedor, J., Garrido, J., Colás, J.: "Use of a hierarchical skeleton for spanish sign language 3d representation over mobile devices"; Proceedings of INTERACCION; 565–568; AIPO, 2006.

López Colino, F.: Sintetizador Paramétrico Multidispositivo de Lengua de Signos Española [Spanish Sign Language Multidevice Parametric Synthesizer]; Ph.D. thesis; Escuela Politécnica Superior. Universidad Autónoma de Madrid (2009).

López-Colino, F., Colás, J.: "Hlsml, una notación de alto nivel para la descripción de mensajes sintéticos en lse"; Proceedings of III Congreso Nacional de Lengua de Signos Española; CNSE - UNED, Madrid, Spain, 2009.

López-Colino, F., Colás, J.: "Hybrid paradigm for spanish sign language synthesis"; Universal Access in the Information Society; y (In press), x–x.

Marshall, I., Sáfár, E.: "A prototype text to british sign language (bsl) translation system"; Proceedings of Annual Meeting on Association for Computational Linguistics; 113–116; Association for Computational Linguistics, Morristown, NJ, USA, 2003.

Morford, J., MacFarlane, J.: "Frequency characteristics of american sign language"; Sign Language Studies; 3 (2003), 2, 213–225.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Herming, J.: HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide; Signum-Verlag, 1989.

Rocha, A., Pereira, G.: "Supporting deaf sign languages in written form on the web"; Proceedings of LREC; 26–28; Lisbon, 2004.

Rodríguez González, M.: Lenguaje de Signos; Fundación CNSE, Barcelona, 1992.

Sagawa, H., Takeuchi, M.: "A teaching system of japanese sign language using sign language recognition and generation"; ACM international conference on Multimedia; 137–145; ACM, New York, NY, USA, 2002.

San Segundo, R., Barra, R., Córdoba, R., D'Haro, L. F., Fernández, F., Ferreiros, J., Lucas, J. M., Macías-Guarasa, J., Montero, J. M., Pardo, J. M.: "Speech to sign language translation system for spanish"; Speech Communication; 50 (2008a), 11-12, 1009–1020.

San Segundo, R., Montero, J. M., Macías-Guarasa, J., Córdoba, R., Ferreiros, J., Pardo, J. M.: "Proposing a speech to gesture translation architecture for spanish deaf people"; Journal of Visual Languages and Computing; 19 (2008b), 5, 523–538.

Sandler, W., Lillo-Martin, D.: Sign Language and Linguistic Universals; Cambridge University Press, 2006.

Schembri, A.: Perspectives on Classifier Constructions in Sign Languages; chapter 1: Rethinking 'classifiers' in Signed Languages, 3–34; Psychology Press, 2003.

Schembri, A., Jones, C., Burnham, D.: "Comparing action gestures and classifier verbs of motion: Evidence from australian sign language, taiwan sign language, and nonsigners' gestures without speech"; Journal of Deaf Studies and Deaf Education; 10 (2005), 3, 272–290.

Schuler, W.: "Using model-theoretic semantic interpretation to guide statistical parsing and word recognition in a spoken language interface"; Proceedings of the 41st Annual Meeting on Association for Computational Linguistics; 529–536; Association for Computational Linguistics, Morristown, NJ, USA, 2003.

Solina, F., Krapez, S., Jaklic, A., Komac, V.: Design and Management of Multimedia Information Systems; chapter 13: Multimedia Dictionary and Synthesis of Sign Language, 268–281; Idea Group Publishing, 2001.

Stokoe, W. C.: "Sign language structure: An outline of the visual communication systems of the american deaf"; Studies in linguistics: Occasional papers; 8 (1960), 1–78.

Supalla, T.: Perspectives on Classifier Constructions in Sign Languages; chapter 11: Revisiting Visual Analogy in ASL Classifier Predicates, 249–257; Psychology Press, 2003.

Sutton, V.: "Signwriting"; http://www.signwriting.org/ (1974).

van Zijl, L., Barker, D.: "South african sign language machine translation system"; Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa; 49–52; ACM, New York, NY, USA, 2003.

VCom3D: "Signsmith studio"; Online (www.vcom3d.com) (2009).

Zwiterslood, I., Verlinden, M., Ros, J., Schoot, S.: "Synthetic signing for the deaf: esign"; Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment; CVHI, Granada, Spain, 2004.