Escuela Politécnica Superior

Departamento de Tecnología Electrónica y de las Comunicaciones

TOWARDS A RULE-BASED SPANISH TO SPANISH SIGN LANGUAGE TRANSLATION: FROM WRITTEN FORMS TO PHONOLOGICAL REPRESENTATIONS

PhD DISSERTATION

Jordi Porta Zamorano

Madrid, November 2014

TESIS DOCTORAL

TÍTULO: Towards a Rule-based Spanish to Spanish Sign Language Translation: From Written Forms to Phonological Representations

AUTOR: Jordi Porta Zamorano

AFILIACIÓN: Human Computer Technology Laboratory
Escuela Politécnica Superior
Universidad Autónoma de Madrid

DIRECTORES: José Colás Pasamontes
Universidad Autónoma de Madrid

Fernando Jesús López Colino
Universidad Autónoma de Madrid

El tribunal nombrado para juzgar la tesis doctoral antes citada, formado por:

PRESIDENTE:

SECRETARIO:

VOCALES:

acuerda otorgarle la calificación de

Madrid, a de de 2014

Universidad Autónoma de Madrid

Escuela Politécnica Superior

Departamento de Ingeniería Electrónica
y de las Comunicaciones

PhD DISSERTATION

TOWARDS A RULE-BASED SPANISH TO SPANISH SIGN LANGUAGE TRANSLATION: FROM WRITTEN FORMS TO PHONOLOGICAL REPRESENTATIONS

Author:
Jordi Porta Zamorano

Supervisors:
José Colás Pasamontes and Fernando Jesús López Colino

Madrid, November 2014

Als meus pares i germans: el Jordi, la Lourdes, l'Albert i l'Anna.

Abstract

This thesis addresses several aspects about the automatic translation from Castilian Spanish to Spanish Sign Language (LSE), two typologically distant languages with not enough linguistics resources enabling statistical approaches to translation. For this reason, a rule-based approach grounded on contrastive grammatical studies on both languages is used.

An architecture following the analysis, transfer and generation model has been chosen. Transfer is performed at the grammatical function level, which is delivered by a Spanish dependency parser without incurring into the complexities of a more deeper analysis.

The bilingual base lexicon is obtained from the *Diccionario normativo de la lengua de signos española* (DILSE-III), which contains the correspondences between Spanish lemmas and their SEA (*Sistema de escritura alfabética*) representation of signs. The lexicon is extended in two different ways: taking advantage of the difference in flexibility between the part-of-speech systems of Spanish and LSE and exploiting several lexical semantic relations, such as synonymy, hyponymy and meronymy.

During the structural transfer phase, some nodes of the dependency analysis are transformed, others are removed and new nodes are inserted. Some classifier predicates are generated in this phase. Surface order generation of signs is obtained by means of the topological ordering of the graph of precedence relations between signs. Pairs of signs having head-dependent relations or sharing the same head are examined in order to determine if its relative ordering is marked or not. The system is evaluated at this point and results are compared to those obtained with statistical models. Best results are obtained with the rule-based approach, with a 0.30 BLEU (Bilingual Evaluation Understudy) and a 42% TER (Translation Error Rate). A linguistic-oriented analysis of errors is provided.

Finally, in the morphological generation phase, glosses with morphological annotations are replaced by the HamNoSys (Hamburg Sign Language Notation System) phonological representations produced by a computational morphology. These representations are used for animation synthesis with avatars. The computational morphology that has been implemented uses inflection, introflection and suppletion to model a significant fragment of the LSE morphology. Among the phenomena considered, it has been implemented deictics, nominal plural, aspect marking, verbal agreement, adjectival modification and degree.

Resumen

Esta tesis aborda varios aspectos sobre traducción automática ed español a lengua de signos española (LSE), dos lenguas tipológicamente distantes y con insuficientes recursos lingüísticos que hagan posible aproximaciones estadísticas a la traducción. Por ese motivo, se propone una estrategia basada en reglas lingüísticas fundamentadas en los estudios gramaticales contrastivos existentes entre ambas lenguas.

Se ha optado por una arquitectura para la traducción siguiendo el modelo de análisis, transferencia y generación, en la que la transferencia se realiza al nivel de las funciones gramaticales proporcionadas por un analizador de dependencias, evitando así las complejidades asociadas a un análisis lingüístico mas profundo para el español.

El lexicón bilíngüe base para la transferencia léxica se ha obtenido de las entradas del *Diccionario normativo de la lengua de signos española* (DILSE-III), que contiene las correspondencias entre lemas en español y la representación SEA (Sistema de escritura alfabética) de los signos. Este lexicón se ha ampliado por dos vías: Aprovechando las diferencias de flexibilidad entre las clase de palabras del español y la LSE, y explotando relaciones semánticas como la sinonimia, la hiperonimia y la meronimia.

Durante la transferencia estructural, algunos nodos del árbol de análisis de dependencias son transformados, otros son borrados y son insertados nuevos nodos. Algunos predicados clasificadores son generados en esta fase. La generación del orden superficial de los signos se obtiene mediante la ordenación topológica del grafo de relaciones de precedencia entre signos. Los pares de signos en nodos que mantienen la relación núcleo-dependiente o son dependientes de un mismo signo son examinados para determinar si su orden relativo está marcado o no. El sistema de traducción es evaluado en este punto utilizando un corpus y comparado con el resultado obtenido con distintos modelos de traducción estadística. Sobre un corpus de control de glosas, el sistema basado en reglas obtiene mejores resultados, con un BLEU (Bilingual Evaluation Understudy) del 0,30 y un TER (Translation Error Rate) del 42%. Sobre los resultados se ha realizado un análisis de los errores.

Finalmente, para la generación morfológica, las glosas junto con sus correspondientes anotaciones morfológicas son reemplazadas por las representaciones fonológicas HamNoSys producidas por una morfología computacional y usables para la síntesis de animaciones mediante avatares. La morfología implementada usa flexión, introflexión y supleción para modelar un fragmento bastante amplio de la LSE. Entre los fenómenos tratados se incluyen la deixis, la realización de los distintos tipos de plural nominal, el aspecto, la concordancia argumental del verbo, la modificación adjetival y el grado.

Acknowledgments

I want to thank my supervisors José Colás and Fernando J. López-Colino, members of the Human-Computer Technology Laboratory (HCTLab) at UAM, for encouraging me to start and continue the work presented in this dissertation. I think I share with José the Humanistic point of view of technology and I quickly was convinced of the importance of the application of language technologies to the field of disability and in particular to deafness. I want to thank him for the many conversations on the way home from the University and for obtaining the economical support to create early prototypes that were shown in forums such as the GSMA Mobile World Congress in 2011. This work complements Fernando's work on Avatars to whom I specially want to thank for all the help I have received in all the spectrum of this work, ranging from the phonological transcription of signs to the bureaucratic procedures around presenting a dissertation. I also want to thank Javier Tejedor, a member of HCTLab for many years, who worked very hard on the data used for experimentation and on the writing of many articles. Thanks also to Marisol Benito, who spokes Spanish and German, and signs LSE and DGS. She not only helped me in solving some doubts about LSE, but in generating many others. Montse Marimon reviewed several drafts of this dissertation and discussed several topics that day in Girona, from the structure and order of the chapters to Linguistic Anthropology, thanks, Montse, also for pointing me out the differences between British and American serial comma. Finally, Marina deserves a mention not only for helping me compiling of the corpus used for experimentation, but for many other things which are not topics of this dissertation.

Contents

List of Tables

List of Figures

List of Morphological Rules

Acronyms

ASL American Sign Language

Auslan Australian Sign Language

BLEU Bilingual Evaluation Understudy

BSL British Sign Language

CNSE *Confederación Nacional de Personas Sordas*

CP Classifier Predicate

DGS German Sign Language

DILSE-III *Diccionario normativo de la lengua de signos española*

DRAE *Diccionario de la Real Academia Española*

EBMT Example-based Machine Translation

EWN EuroWordNet

FST Finite-State Transducers

GDLSE *Gramática didáctica de la lengua española*

HPSG Head-driven Phrase Structure Grammar

HamNoSys Hamburg Sign Language Notation System

INE *Instituto Nacional de Estadística*

IPA International Phonetic Alphabet

IS International Sign

LAS Labelled Attachment Score

LCM Labelled Complete Match

LSC Catalan Sign Language

LSE Spanish Sign Language

MECD *Ministerio de Educación, Cultura y Deporte*

MT Machine Translation

NGT Sign Language of the Netherlands

NMC Non-manual Component

PER Position Independent Word Error Rate

PoS Part of Speech

POSIX Portable Operating System Interface

RBMT Rule-based Machine Translation

SEA *Sistema de escritura alfabética*

SES Signed Exact Spanish

SiGML Signing Gesture Markup Language

SL Sign Language

SMT Statistical Machine Translation

SOV Subject-Object-Verb

SVO Subject-Verb-Object

TER Translation Error Rate

WER Word Error Rate

XML Extensible Markup Language

Chapter 1

Introduction

Translation helps people to communicate across linguistic and cultural barriers. However, according to Isabelle and Foster (2005), translation is too expensive and its cost is unlikely to fall substantially enough to constitute it as a practical solution to the everyday needs of ordinary people. Although it remains to be seen if machines will ultimately compete seriously with humans in translation, machine translation can help break linguistic barriers and can make translation affordable to many people. This situation is especially important for Deaf people[1], since translation helps Deaf and Hearing communities to communicate with each other and provides Deaf people with the same opportunities to access information as everyone else. Technology had always been played an important role in the integration of the Deaf into general society. This dissertation aims at making some contributions to the automatic translation of written Castilian Spanish into phonological transcriptions of Standard Spanish Sign Language, which can be used to generate animations using the existing avatar technology.

1.1 Motivation

In 1999/2000, according to INE[2] and MECD[3], there were 1,064,000 people in Spain with hearing problems. Half of that population was more than 65 years old. Among the Spanish deaf population, 47% of them did not have basic-level education and in many cases they were illiterate. Only one to three percent of Spanish deaf people had graduated from university (INE, 1999; MECD, 2000). Moreover, about 92% of the Spanish deaf population had significant difficulties in using written Spanish (INE, 2008). The unemployment

[1]The distinction between 'deaf', to refer to the physical condition of deafness, and 'Deaf', to refer to the deaf culture, has been widely adopted.

[2]*Instituto Nacional de Estadística* (Spanish National Statistics Institute)

[3]*Ministerio de Educación, Cultura y Deporte* (Spanish Ministry of Education)

rate for deaf people was almost 20% and that of deaf women was as high as 30% (INE, 2003).

Spanish Sign Language (LSE) is the name of the most widespread gestural-visual language used by the Spanish Deaf community and by persons who live and interact with them. According to Amate García (2001), the estimated number of users of LSE was 100,000. For about 20 to 30% of those users, LSE is their second language. The ratio of interpreters to LSE users in Spain is 1/221, below the European average, which is 1/10.

From a linguistic point of view, LSE refers to a variety of the sign language used in a large central area of Spain, with Madrid being its cultural and linguistic epicentre. There are other local varieties in some areas of Asturias, Aragón, Murcia, Western Andalucía (Seville), etc. The mutual intelligibility among these varieties is high due to their lexical similarities (Parkhurst and Parkhurst, 2001). LSE was finally recognised as an Spanish Language in 2007 (*LEY 27/2007*).

1.2 Myths and Truths on Sign Languages

Sign languages exploit a different physical medium from the oral-aural system of spoken languages. Sign languages are gestural-visual languages. This difference in modality causes sign languages to constitute another branch within the typology of languages. However, there are still many myths around sign languages. One of the most common and enduring myths is that sign language is universal[4]; however, every Deaf community has its own sign language, even within the same country. For example, in Spain, apart from Spanish Sign Language (LSE), there exists another recognised sign language, known as Catalan Sign Language (LSC) and used within the Catalan Deaf community. Another common myth is that there is a correlation between spoken and sign language families. American Sign Language (ASL) and British Sign Language (BSL), however, despite the fact that both are sign languages used in English-speaking countries, are mutually unintelligible. Sign languages do not derive from spoken languages, but, as any other languages, can be influenced by contact with other languages. As with spoken languages, when the use of an sign language is extended, dialects and varieties are developed. Sign languages are natural languages that arise spontaneously in Deaf communities to fulfil the function of communication and they possess the features that characterise human languages (Hockett, 1960).

Phonocentrism is a view in which speech is considered to be superior to, or more natural than, written language. This attitude, which is still dominant in western culture, has

[4]International Sign (IS), sometimes referred as Gestuno (Rubino, Hayhurst and Guejlman, 1975), is seen as a pidgin form of sign language, simpler than a sign language and with a limited lexicon.

negatively affected the consideration of sign languages, adding to their status of minority languages the status of minorised languages, i.e., languages whose value is not recognized on the interactional scene by speakers of a sociolinguistically dominant language. This also encourages the assumption that speakers of the minorised language conform to the usage and interactive norms set by their interlocutors. Nevertheless, scientific claims regarding the status of sign languages as 'real' human languages have been made since the work of W. Stokoe on the ASL (Stokoe, 1960). Sign languages in developed countries, especially sign languages in Europe and North America, dominated research during the first decades of study. Currently, language typologists still have some difficulties accessing research on a number of regions like Central and South America, Africa or Asia because most publications are written in national languages not accessible to a wider international audience. According to Zeshan (2007), the state of knowledge regarding sign languages has developed like a mosaic with many untiled gaps, but this is increasingly giving typologists a clearer picture of the range of diversity in sign languages. Some cross-linguistic and typological studies of sign languages, such as those by Sandler and Lillo-Martin (2006), Brentari (2010) or Pfau, Steinbach and Woll (2012), have shed light on both the universals and the diversity of sign languages, making valuable contributions to the understanding of human languages in general.

The lack of a writing system is a characteristic of sign languages and is shared with two-thirds of the spoken languages of the world. Strictly speaking, the only way of representing sign languages is to use motion pictures. However, several notational systems have been developed. The most important ones today are SignWriting (Sutton, 1981) and HamNoSys (Hamburg Sign Language Notation System; Prillwitz et al. (1989)). SignWriting was conceived primarily as a writing system and has its roots in DanceWriting (Sutton, 1973), a notation for reading and writing dance movements. HamNoSys was conceived as a phonological transcription system for sign languages, with the same objective as the International Phonetic Alphabet (IPA) for spoken languages (International Phonetic Association, 1999). There is another alphabetic writing system, designed specifically for LSE, called SEA (*Sistema de escritura alfabética*; Herrero Blanco, Alfaro and Cascales (2001)); this one uses the Latin alphabet and has LSE's phonology as its basis. However, when syntax is under the focus, glosses replace the phonological representations of signs. Glossing is a commonly used system for explaining or representing the meaning of signs and the grammatical structure of signed phrases and sentences in a text written in another language. However, glossing is not a writing system that could be understood by sign language users. A machine translation system needs to produce an animation to be considered a complete and useful system.

Most contemporary works on sign languages have adopted language theories created

for spoken language instead of developing new theories. This adoption leads naturally not only to the study of the phonology, morphology and morpho-syntax of sign languages, but also to the study of all other descriptive levels found in spoken languages. However, from the point of view of natural language processing, Sign languages are still under-resourced or low-density languages – that is to say, little or no specific technology is available for these languages and computerised linguistic resources, such as corpora or lexicons, are very scarce. This situation, of course, is not exclusive to sign languages, since it in fact applies to most of the languages of the world.

1.3 Contributions and Organisation of the Dissertation

The main contributions in this dissertation to the automatic translation of Castilian Spanish to Standard Spanish Sign Language can be summarised in the following key points:

- Architecture: Since LSE is a low-density language, a linguistic approach based on contrastive studies between Castilian Spanish and Standard LSE is proposed as a suitable alternative to data-oriented approaches. The level of functional dependencies has been identified as the appropriate level for a linguistic transfer-based machine translation.

- Bilingual resources: The initial bilingual lexicon has been increased in size by using Spanish morpho-lexical relationships and exploiting the differences in part-of-speech flexibility beetwen Spanish and LSE. In addition, lexical-semantic relationships have been used to bridge the lexical gap and to relate particular groups of words and their translational equivalences as classifier names.

- Word order generation: An algorithm has been designed to deal with the topic-oriented surface order of LSE. The algorithm makes use of linear precedence rules operating at the level of syntactic functions and dealing with topicalisation.

- Evaluation and comparison with statistical approaches: A parallel Spanish-LSE glosses corpus is created from the data used in a psycholinguistic study on LSE (Rodríguez Ortiz, 2005). A subset of this corpora has been used to evaluate the system. Experiments reported a BLEU (Bilingual Evaluation Understudy) about 0.30 and a TER (Translation Error Rate) at about 42%. Results are compared with those obtained using phrase-based statistical machine translation with different settings. A linguistic-oriented error analysis shows that many of the differences between the system output and the reference translations arise from variations in the realisation

and grom the doubling of the linguistic structures. In addition, the analysis shows that classifier predicates, which are one of the cornerstones of sign languages, are the most difficult expressions to generate. However, they have a translatable subclass corresponding to Spanish prepositional expressions with locative and temporal meanings.

- LSE morphological generation: A finite-state computational inflectional morphology is described for a significant fragment of LSE non-linear morphology. Finite-state transducers are used to alter the phonological base forms of signs to express grammatical values by means of affixation, introflection, suppletion and reduplication. Phenomena considered are deixis, nominal plurals, verbal agreement, aspectual marking, adjective modification and degree. In the generation phase, this component produces a phonological description that can be synthesised by avatar technology.

Most of the results of this dissertation have been published in:

- López-Colino, Fernando J., Jordi Porta and José Colás. 2010. Linguistic principles applied to Spanish Sign Language automatic synthesis. *Estudios de Lingüística Universidad de Alicante (EULA)*, (24):167–198.

- Porta, Jordi, Fernando J. López-Colino and José Colás. 2012. A computational morphology for Spanish Sign Language nominal inflection. In *Proceedings of IberSpeech-2012*, Madrid, Spain, November.

- Porta, Jordi, Fernando J. López-Colino, Javier Tejedor and José Colás. 2014. A rule-based translation from written Spanish to Spanish Sign Language glosses. *Computer Speech & Language*, 28(3):788–811.

Also some prototypes of the machine translation system have been publised in:

- López-Colino, Fernando J., Javier Tejedor, Jordi Porta and José Colás. 2011. Integration of a Spanish-to-LSE machine translation system into an e-learning platform. In *Proceedings of the 6th International Conference on Universal Access in Human-Computer Interaction. (UAHCI-2011)*, pages 567–576, Orlando, USA.

- Tejedor, Javier, Fernando J. López-Colino, Jordi Porta and José Colás. 2012. An on-line, cloud-based Spanish-Spanish Sign Language translation system. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (InterSpeech-2012)*, pages 2125–2126, Portland, Oregon, USA.

- Porta, Jordi, Fernando J. López-Colino, Javier Tejedor and José Colás. 2012. An on-line system adding subtitles and sign language to Spanish audio-visual content. In *Proceedings of the IberSpeech-2012*, Madrid, Spain, November.

This dissertation is organised as follows:

- Chapter 1 introduces this work by presenting the motivation and placing Sing Languages within the diversity of human languages according to its typological and sociolinguistic characteristics.
- Chapter 2 introduces a grammatical sketch of LSE describing the most salient linguistic phenomena of that language without going into deeper details, but giving a background to the rest of the work. Many of the phenomena described are shared with other sign languages.
- Chapter 3 presents the different stages in the transfer-based architecture of the machine translation system and the creation of the transfer lexicon. This Chapter also contains a review of the previous approaches to the translation into sign languages, the existing sign language Corpora and an evaluation of the system proposed using a parallel corpus.
- Chapter 4 contains a computational morphology of LSE. The range of the phenomena implemented justifies a separate chapter devoted to this subject.
- Chapter 5 contains the conclusions and the future work on the topic of this dissertation.

Chapter 2

A Grammatical Sketch of Spanish Sign Language

The most important documentation for a language is a reference grammar, which documents the principles governing the construction of words and all kinds of grammatical structures found in a language. The first Spanish Sign Language (LSE) grammar has been published in *Gramática didáctica de la lengua de signos española* (GDLSE) (Herrero Blanco, 2009). It is a complete grammar, describing the linguistic structures of LSE with a pedagogical orientation. Since the grammar is addressed to students and teachers of LSE, it includes exercises and multimedia material. Furthermore, the *Diccionario normativo de la LSE* (DILSE-III) (Fundación CNSE, 2008) represents another milestone for LSE. Despite being descriptive works, both works have a normative-oriented guidance, so it is in their aim to represent the contemporary standard for LSE.

2.1 Phonology and Non-manual Components

Early studies on sign languages used the term 'cherology', but now sign language linguistics prefer the term 'phonology' to emphasize that this sublexical level is organisationally and functionally equivalent to that in oral languages despite differences in modality (Nöth, 1990).

The articulators in sign languages are the hands, the arms, the face (including eyes, eyebrows, mouth, cheeks and tongue), the head, the neck, the shoulders and the body. During signing, the visual field of one signer covers all the articulatory elements, hands and body, of the other signer. However, in a conversation, attention is focused on the gaze of the other signer, leaving the rest in the peripheral vision. The gaze plays a major role in sign language communication, as in spoken communication.

In LSE and in other sign languages, hands are the main articulator. The dominant

hand is the hand involved in the articulation of one-handed signs. In two-handed signs, each hand can act either symmetrically or antisymmetrically, or there can be a dominant hand acting as a coordinator and a so-called non-dominant hand, or passive hand, used as the place of articulation. In right-handed signers, the dominant hand is the right hand and the left hand is the non-dominant hand. The role of the hands is reversed in the case of left-handed signers. This difference, however, does not cause confusion in understanding.

Signs are formed by a set of minimal, meaningless elements or parameters, whose variation expresses the appearance of a new sign. There are four main kinds of these articulatory parameters or phonemes: the handshape, the hand orientation, the place of articulation and the movement. Moreover, the contact of the hand with the body or the other hand, and the type of symmetry in two-handed signs, are also considered phonologically relevant in LSE. Some studies also consider the non-manual components (NMCs) of a sign as parameters of that sign, but the number of signs whose meaning depends on NMCs in LSE is so few that this is usually ignored in phonology. NMCs sometimes play a role similar to the tone in spoken languages.

NMCs have several functions in morphology, and can be performed alone or in conjunction with the manual component. In LSE, the most significant functions are the expression of intensification or restriction, and the expression of the verbal modality. Intensification/restriction is usually applied to adjectives, and can be alternatively expressed in LSE by means of adverbs as 'MUCHO (much)' or 'ALGO (some)'. In the case of the non-manual expression of intensification, clenched teeth, semi-closed eyes and tense articulation denote the intensification of a positive property, as for 'ALTÍSIMO (very tall)', and intensive negative properties are denoted by inflated cheeks and blowing, as for 'MUY-CERCA (very close)'. These NMCs can be also applied to classifiers representing degree in adjectives. Some verbal aspects have no manual realisation. This is the case with the inchoative aspect value, for which the performance of the sign is suspended, with open mouth and raised eyebrows. Deontic modality is marked with raised eyebrows and pursed lips, whereas epistemic modality uses frowns and a U-like mouth shape, along with an affirmative movement of the head. The main functions of NMCs in syntax are the distribution of information and agreement. Topicalised information is marked with pauses and eyebrows. Agreement can be marked by the gaze and directing the body towards the position of the space where the object or the subject of a verb has been previously placed. The interrogative construction in LSE is realised by raising the eyebrows and bending the torso forward while signing the verb of the sentence. The negative constructions may be realised by including the sign 'NO (not)' after signing the verb. However, it is also common simply to negate with the head while signing the verb. NMCs also have other roles in discourse and in conversation (in particular in turn-taking), but these fall outside

the scope of this work.

2.2 The Use of Space

Signs are produced in the signing space, i.e., the physical space around the signer. The space used for the articulation of signs is a lexical space and has a phonological value. This space comprises some body areas and the central and lateral planes of symmetry. However, space has additional and complex uses. Space can be used topographically, i.e., as a map, when visible objects used in communication are signalled, or analogically, when objects are placed in analogy to the real world, this use being iconic or metaphoric. Some extensions of the frontal lexical space have grammatical purposes. They are used for deixis, anaphora and agreement. Space is used syntactically, assigning arbitrary places (the loci) for pronominal reference. These locations have no relation to the physical locations of the objects situated there, since these objects can be abstract. In subsequent references, grammatical relationships to objects can be established through locations, which can act as morphemes.

2.3 Parts of Speech

Hengeveld, Rijkhoff and Siewierska (2004) distinguish among specialised, flexible and rigid languages, depending on how grammatical functions and lexemes relate to each other. In specialised or rigid languages, each function has a particular lexeme, whereas flexible languages display classes of lexemes that can be used in more than one function without requiring lexical or syntactic derivation. Flexibility generates homography. Spanish is considered a specialised language but, as Herrero Blanco (2009) points out, LSE displays some flexibility in its part-of-speech (PoS) system. In LSE, sometimes the noun, the adjective and the verb are represented by the same sign, as is the case for '*dolor* (pain)', '*doloroso* (painful)' and '*doler* (to hurt)', which are all represented by the single sign DOLOR.

2.4 Classifiers

According to Aikhenvald (2000), almost all languages have some grammatical means for the linguistic categorisation of nouns and nominals. Classification is a categorisation device found in languages for marking humanness, animacy, shape, size and other inherent and salient properties of objects. In sign languages, the phenomenon of classification is pervasive and has received different analyses in the sign language literature (Schembri,

2003). Supalla (1986) defined classifiers as morphemes incorporated into verbs, but Sutton-Spence and Woll (1999) imposed three conditions on classifiers: to reference a group of elements sharing some common features, to be a proform and to be found in verbs of movement and location. Proforms have an anaphoric primary function (Engberg-Pedersen, 1989) and can be defined as anything that refers to, and stands in the place of, something previously presented in the discourse. Consequently, pronouns are a subclass of proforms. Proforms are realised as a handshape with a specific orientation, which represents the prototypical orientation of the object. Both parameters can be considered morphological components of the proform. For example, a proform can represent a person, using a particular handshape with a horizontal orientation to indicate that the person is lying down.

Broadly speaking, we will refer to classifiers as handshapes that substitute for other signs and have a morphological value. There are three groups of classifiers in LSE: entity classifiers, handle classifiers and extension classifiers. Entity classifiers are descriptive classifiers that represent objects by their predominant dimension or shape. Therefore, an object can be classified according to its dominant dimension as one-dimensional (pen, spoon, etc.), two-dimensional (book, sheet, table, etc.), circular (dish, coin, etc.), three-dimensional (ball, building, fruit, etc.), cylindrical (bottle, column, etc.) or fluid (water, smoke, crowds, etc.). Four-wheeled and two-wheeled vehicles have different classifiers, as do have objects with a framework (swimming pool, picture, etc.). Handshapes for descriptive and entity classifiers can be seen in Table 2.1. Other classifiers can represent the thickness of an object or stacked objects. Persons, animals and parts of the body also have classifiers. Handle or instrument classifiers describe the handling of an entity rather than representing the entity itself. Extension classifiers depict the perimeters of objects or make reference to their surfaces. Classifiers, as well as path movements, are among the most iconic elements of sign languages (Supalla, 1986).

2.5 Morphology

Grammatical values can be expressed in the form of independent single words, called free morphemes, or as additions or alterations to the phonology of words, in which case they are called bound morphemes. Not all languages have the same grammatical values or express them in a similar way. For example, Spanish has gender morphemes, but LSE expresses gender values by other means, using free morphemes. Languages can be classified according to the degree of presence of the two kinds of morphemes. At one end of the spectrum there are isolating languages, like Vietnamese, with a low morpheme-per-word ratio, and at the opposite end there are synthetic languages, like Inuit languages,

with lengthy words. However, in general, the vast majority of languages, including sign languages, have both kinds of morphemes. In addition, sign languages have a widely used morphological resource for the phenomenon of classification noted in Section 2.4.

Sign languages have not only inflection but also derivation, compounding and classification (see Section 2.6). Inflectional morphology is the study of the changes undergone by words in expressing grammatical values. LSE, and virtually all sign languages studied (Aronoff et al., 2004), have a rich inflection system. Bound morphemes found in LSE belong to one of two classes, flective or introflective. Flective morphemes are concatenated to the phonological form, like a movement or a repetition[1] of the original movement, sometimes repeating in an specific direction. They are mainly used for expressing several forms of plural, some values of the verbal aspect, or as a mechanism for morphological derivation. Introflective morphemes cause internal changes to the phonological form of the sign, and hence they are non-concatenative in nature. This kind of morpheme is also found in Semitic languages. Introflection is used to agree with the subject, the object, the receiver or the place. The classifier of the subject and/or the object are also introflected in verbs. The number is introflected in pronouns, etc. LSE has, in some cases, irregularities such as suppletive base forms, i.e., different unrelated base forms for the same sign, indicating different grammatical values.

LSE does not mark gender in signs, and hence gender is expressed by adding 'HOMBRE (man)' / 'MUJER (woman)' after the sign. Nouns and adjectives can inflect, changing some of their phonological components to express number, to be classified, to express place or agreement with a subject, etc. Nouns admit adjectival inflection of size and shape. There are several ways of forming plurals in LSE, depending partly on the phonological structure of signs. Personal pronouns inflect in number and, like some temporal nouns, they can incorporate numerals. Place deictics have values for first, second and third person.

Additionally, in LSE, the third person deictics and personal pronouns are split into two values: proximal or present, and distant or absent. Adverbs can be modified to express intensification or restriction by means of the execution speed and accompanying facial and other non-manual components. Verbs can inflect for both subject and object agreements, and can express a variety of aspectual values such as habitual, continuous, etc. Verbal agreement in LSE is not as systematic as in Spanish, where all verbs agree in number. Padden (1990) proposed a classification for verbs of several sign languages on the basis of which affixes may be added. She proposed three basic classes: plain, agreement and spatial verbs. Plain verbs neither inflect for person and number nor take locative affixes. Agreement verbs use the syntactic space and can inflect in person and number,

[1] Repetition can be seen as reduplication, another non-concatenative operation found in some languages as Tagalog.

but cannot incorporate spatial information. Finally, spatial verbs use the topographical space and incorporate spatial information, but cannot incorporate person and number. Aspect marking is widely used by sign languages. In LSE, reduplicating a sign rapidly can express habitual aspect. An abrupt stop before the completion of the sign marks an unrealised inceptive aspect. If the sign is repeated with a slower circular movement, its aspect value is continuous. Some nouns and adjectives can undergo aspect marking, for instance with habitual values, e.g., 'ENFERMO (sick)' > 'ENFERMIZO (sickly)', and continuous (monotonous) values. All the classes of verbs can be slightly modified to incorporate aspectual values.

2.6 Classifier Nouns

A classifier noun is a common noun acting as a free morpheme prepended to another noun to complement its lexical meaning, or to other lexical categories to produce lexical derivatives. For instance, LSE resorts to classifier nouns for tree names, so that the Spanish noun '*naranjo* (orange tree)' has a translational equivalence as a construction headed with the sign 'ÁRBOL (tree)' and followed by the sign for the fruit of the tree 'NARANJA (orange)' forming 'ÁRBOL NARANJA (lit. tree orange)'. The deverbal noun '*trabajador* (worker)' has an equivalence with 'PERSONA TRABAJAR (lit. person to.work)' and the adjective '*envidioso* (envious)' as 'PERSONA ENVIDIA (lit. person envy)'. For LSE, the most widely used classifier nouns are 'AGUA (water)', 'ÁRBOL (tree)', 'CASA (house)', 'DINERO (money)', 'TRABAJO (work)', 'PERSONA (person)' and 'TIENDA (shop)'.

According to Herrero Blanco (2009), classifier nouns present a different degree of grammaticalisation so that some of them can be suppressed, as is the case with 'PERSONA (person)', but not with 'TIENDA (shop)' and, in general, they do not need to be realised when they have been used before or when the context contains enough information to infer them. More examples of Spanish words that have translational equivalences with classifier noun constructions are '*cuartel* (barracks)' → 'CASA MILITAR (lit. house military)', demonyms (the name given to a person from a particular region or country), like '*europeo* (european)' → 'PERSONA EUROPA (lit. person Europe)', '*billete* (bill)' → 'DINERO TARJETA (lit. money card)' and '*apicultura* (beekeeping)' → 'TRABAJO ABEJA (lit. work bee)'.

2.7 Classifier Predicates

Classifier predicates pose a challenge to the definition of what constitutes a linguistic expression. Frequently, classifier predicates are spatial metaphors, scene visualisations, or

even pantomimic descriptions. A classifier predicate is a complex sign typically preceded by a nominal phrase. A classifier predicate is created by selecting a classifier. The signer performs a three-dimensional movement with the hand that communicates an outline, a position in the space around the signer, a movement through the dimensional space, a physical or abstract dimension, or any other property of the object that needs to be communicated. This type of predicate is ideal for describing scenes, manipulation tools, movements, size and other information of visual or spatial scenes or processes.

2.8 Syntax

Traditionally, languages have been classified according to the prevailing order of syntactic functions (Greenberg, 1963). For example, both Spanish and English have a Subject-Verb-Object (SVO) basic order, although the Spanish word order is freer than that of English due to the richer morphology of Spanish. Other typological studies have highlighted the fact that some languages do not organise the structure of the basic components of the sentence from the syntactic functions of subject and object but from the grammaticalisation of the functions of topic and comment, as do Mandarin, Mayan languages and the languages of the Philippines (Kiss, 1995). However, still other studies (Li and Thompson, 1976) have shown that some languages use both types of organisation, resulting in four types of languages: subject-prominent languages, topic-prominent languages, languages prominent in either subject or topic, and languages organised according to other parameters.

Early research in ASL word order during the 60's was influenced by the functional-oriented typology of that decade. In the 80's, the description of other sign languages led other researchers to consider other motivations, such as semantics or pragmatics, to explain the word order of different sign languages (Leeson and Saeed, 2012). Sign languages have a considerable variation of basic word order of a sort very similar to that found in the so-called discourse-oriented languages, including Chinese and Russian. Sign languages productively employ devices for presenting information in a different order determined by discourse factors, such as presenting the old information (topic) before the new (comment) and specific non-manual markers are present when the order is other than the basic. Topicalisation, i.e., movement of the topic element at the beginning of the sentence, is widespread in sign languages, but to varying degrees.

In general, LSE is considered a topic-prominent language where topicalisation occurs only in the main clause. The topic is usually marked by NMCs and followed by an intonation break.

Object Type	Handshape	HamNoSys	Examples
One-dimensional			knife, spon, stick, cigarette
Two-dimensional			sheet, book, table
Three-dimensional			ball, house, orange
Circular			
Small			coin, button
Big			wheel, dish, chesse
Cylindrical			bottle, cup, column
Fluids			water, smoke, crowd
Vehicles			
Two-wheeled			motorbike, bicycle
Four-wheeled			car, truck
Framed Objects			picture, swimming pool
Persons			
Single Person			anthony, mary
Two Persons			Mary and Anthony
Many Persons			crowd
Moving Person			walker, runner
Animals			
Bipeds			person, Anthony
Quadrupeds			dog, horse
Birds and Reptiles			bird, lizard
Body Parts			
Head			head
Legs			legs
Feet			feet

Table 2.1: Handshapes for descriptive and entity classifiers in LSE and their corresponding HamNoSys (Hamburg Sign Language Notation System) transcriptions. Drawings in handshape column contain also SignWriting symbols.

Chapter 3

A Rule-based Spanish-to-LSE Translation System

The linguistic rule-based approach to translation is mainly motivated by the fact that available resources for data-driven approaches, as will be seen in Section 3.1, are especially costly to acquire and it is difficult for data-driven systems to estimate parameters from small corpora and still have reasonable coverage in wider domains. The source of knowledge for rule-based machine translation is contrastive linguistic studies. Contrastive linguistics focuses on similarities and differences in the structure and functioning of two or more languages. For the case of Spanish and Spanish Sign Language (LSE), contrastive studies can be found in by Minguet Soto (2000) and in *Gramática contrastiva español / LSE*, published as an electronic resource at the *Biblioteca Virtual Miguel de Cervantes*[1]. Additionally, some other, more specific contrastive studies, such as that of Villameriel (2008), can be found for discourse markers and other linguistic phenomena occurring in discourse or conversations, such as role shifting or turn-taking.

Figure 3.1 shows the architecture of the system presented here and its connection to other external modules: an automatic speech recognizer and an animation synthesiser. It is a classic transfer-based architecture (Vauquois, 1968) where, in the case of Spanish into LSE translation, the transfer phase is effectuated mainly at the level of syntactic functions from dependency analyses. A dependency tree from the analysis module abstracts the constituency structure and the surface word order in the source sentence. Only functional relations between words are represented in a dependency tree. The transfer module makes use of the bilingual lexicon and the knowledge in the language pair-specific rule database to map dependency structures and to transfer a Spanish dependency tree into the corresponding dependency tree in LSE. Word ordering and morphological rules are applied to the transferred dependency tree so that the output of the generation stage is

[1] <http://bib.cervantesvirtual.com/seccion/signos> (Accessed: November 2014)

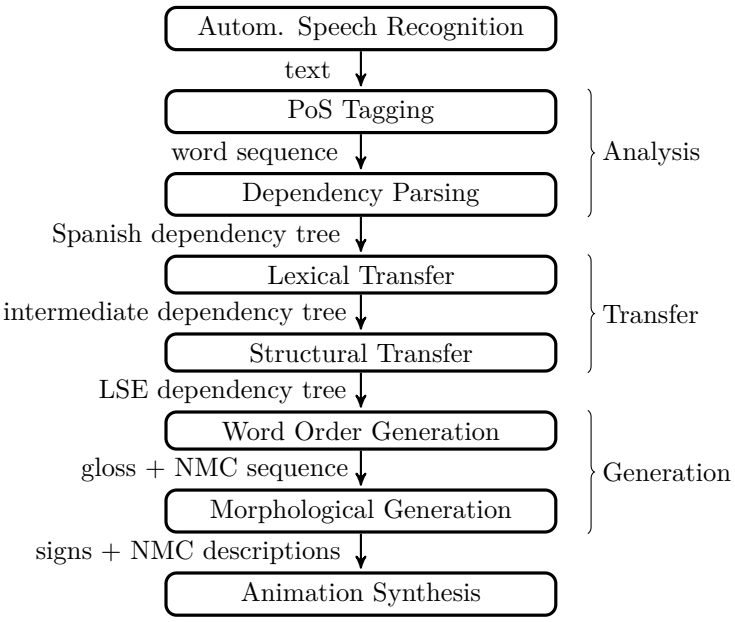


Figure 3.1: Architecture of the Spanish-to-LSE Machine Translation System.

a sequence of glosses with morphological indications. Glosses should be considered an intermediate result from which the corresponding phonological forms are produced by the morphological generation module. However, it is important to note here that these phonological forms can not be considered a true translation, since they should be performed in order to be considered signs. The different stages of the machine translation strategy are detailed next, after reviewing other systems and corpora.

3.1 Review of Machine Translation Systems to Sign Languages and Sign Language Corpora

As was reported by Huenerfauth (2003), all machine translation systems for sign languages (SLs) published up to 2003 were just works in progress or simple demonstrators. However, some systems were particularly distinguished, including the ZARDOZ system (Veale and Conway, 1994; Veale, Collins and Conway, 1998), the ViSiCAST Translator (Bangham et al., 2000; Sáfár and Marshall, 2002), which had subsequent publications (Marshall and Sáfár, 2004), the ASL Workbench (Speers, 2001) and the TEAM project (Zhao et al., 2000). All these systems were rule-based, following the predominant machine translation model of that time and made use of either transfer- or interlingua-based approaches.

The only approach dealing with the translation of classifier predicates was that of Huenerfauth (2006) for English-ASL, who presented a multi-path approach combining interlingua, transfer and direct approaches as a whole. His approach makes use of a 3D

Corpus	Languages pair	Sents.	Tokens		Types		Reference
			Words	Signs	Words	Signs	
-	Italian-Italian SL	585	15,000	6,000	1,442	300	(Bertoldi et al., 2010)
RWTH-Phoenix	German-German SL	2,468	16,500	10,500	1,302	1,895	(Bungeroth et al., 2006)
-	Chinese-Thaiwanese SL	1,983	-	11,501	-	2,159	(Chiu et al., 2007)
-	Catalan-Catalan SL	199	2,416	4,305	446	648	(Massó and Badia, 2010)
ATIS	English-Irish SL	595	4,436	4,333	600	544	(Bungeroth et al., 2008)
ATIS	German-German SL	595	4,903	4,291	627	498	”
ATIS	English-South African SL	595	4,436	2,525	600	422	”
ID	Spanish-Spanish SL	2,000	20,000	15,000	800	400	(San-Segundo et al., 2010)
DL	Spanish-Spanish SL	2,000	11,000	8,000	1,000	500	”

Table 3.1: Counts for sign language parallel corpora for different language pairs.

scene generator from natural language descriptions. These 3D scenes, together with a set of predefined templates, were used to define the position of the hands and perform the classifier predicate.

Today's prevalent research approaches for general machine translation are data-driven. Example-based machine translation (EBMT) and statistical machine translation (SMT) have replaced the earlier rule-based machine translation (RBMT) approaches. Results obtained using these EBMT and SMT approaches have reached such a high level of quality that they make machine translation genuinely useful, especially when the target language is English. However, data-driven approaches estimate their parameters from an aligned bilingual corpus and their accuracy depends heavily on the quality and size of this corpus. Unfortunately, corpora for sign languages are still far from reaching the state of the art of those for spoken languages. The problem of modality and the lack of a standardised writing system make data acquisition for sign languages a time-consuming and expensive task compared to the acquisition of spoken or written data. Sign language data are usually acquired first by video-recording the signer's performance and then aligning the video sequence with multi-channel annotations that describe in a gloss format the signs and the non-manual components. The use of annotation tools such as ANVIL (Kipp, 2001), SignStream (Neidle, 2002), iLex (Hanke, 2002), ELAN (Wittenburg et al., 2002), or the annotation tool presented by Braffort et al. (2004) makes easy the construction and annotation of new corpora for sign languages. On the whole, if corpora are essential for language research, parallel corpora are essential for machine translation. To construct machine translation corpora, sign language data annotations are extracted and aligned with the equivalent expressions in other languages. The amount of parallel data currently available for sign languages is typically from few hundred to some thousand sentences or even just isolated words in restricted domains and on very few topics. However, some parallel corpora are being used for research in machine translation; these are summarised in Table 3.1.

Despite the lack of parallel corpora, the success of data-driven approaches to machine

translation between spoken languages has led to the application of the same techniques to sign languages. However, according to Morrissey (2011), most research in sign language machine translation, with a few exceptions, has emanated from sporadic and short-term projects as opposed to long-term research investment. Some works are still worth mentioning: The Thai-to-Thai Sign Language machine translation system (Dangsaart et al., 2008) presents a direct translation system with reordering rules. The system for Thai reaches an F-score of about 97% for a set of 297 test sentences. Morrissey (2008) presented exhaustive experiments on the MaTrEx, a hybrid approach combining EBMT and SMT (Stroppa and Way, 2006). Results of MaTrEx on the ATIS corpus were about 0.39 BLEU (Bilingual Evaluation Understudy) for English-to-Irish Sign Language translation and about 50% F-score for German-to-German Sign Language (DGS) translation.

For Spanish to LSE, Baldassarri and Royo-Santas (2009) described a rule-based demonstrator. Spanish is analysed using FreeLing (Atserias et al., 2006) dependency analysis, in which the relations between words are interpreted as the relations between blocks, such as subject, predicate, verb, etc. Through the application of a series of grammatical rules, the dependency analysis is transformed into a series of glosses. These grammatical transformations are carried out in reverse order. During transformation, some information flows in a bottom-up direction: partial translation and semantic information. Semantic information includes meaning, temporal and numeral information and information regarding the type of block. A morphological transformation is applied to glosses before generating signs. In this stage, some glosses are replaced by their synonyms when they do not correspond to signs, or by their hyponyms when they have neither equivalent sign nor synonyms. In addition, some individual glosses can be replaced by sequences of signs. All these replacements are based on a dictionary of equivalences. The system was tested with 92 sentences containing a total of 561 words. Sentences varied in length, ranging from simple phrases to more complex ones containing up to 25 words and several verbs along with subordination structures. Appropriate dictionary entries were created for the evaluation, with very satisfactory results: 96% of the words were correctly translated and 93.7% of them were in correct order.

It is also worth mentioning the existence of another rule-based Spanish-to-LSE machine translation system based in Apertium, a free/open-source platform (Forcada et al., 2011). The system uses shallow transfer rules and delivers an SEA (*Sistema de escritura alfabética*) representation of LSE. To our knowledge, except for the work of Gómez-Alcaraz (2011), there are no published results on this system, but it is available on the Internet.[2]

The most complete work in Spanish-to-LSE translation to date is summarized in San-Segundo et al. (2011). Their system focuses on helping Deaf people in a highly specific

[2] <`http://aplica.prompsit.com/en/text_es_ssp`> (Accessed: November 2014)

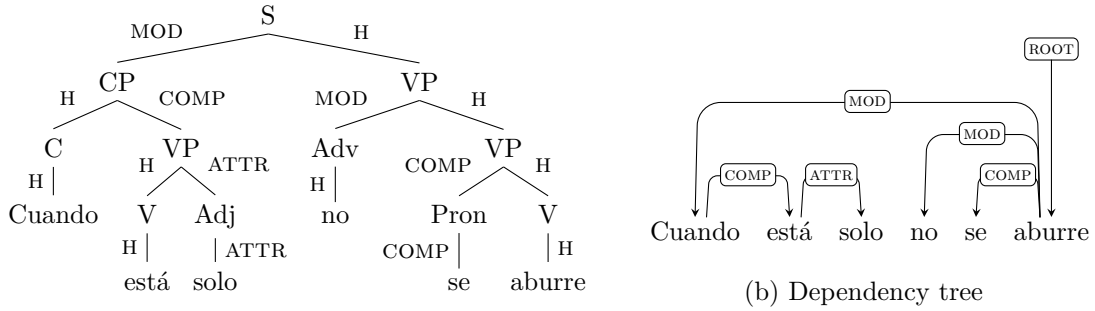
domain, the renewal of driving licenses. The final version of the system combines three approaches in the following order of decreasing priority: an EBMT strategy, a rule-based translation method and an SMT approach. All three approaches compute a confidence value and thresholds determine when to use the next approach. EBMT uses a set of translated sentences and a heuristic distance function based on the Levenshtein distance (Levenshtein, 1966). In the rule-based approach, every word is first mapped into syntactic-pragmatic categories. After that, handcrafted rules are applied bottom-up to convert tagged words into signs. Rules define short- and long-scope relationships between concepts or signs. Categorisation, as described in López-Ludeña et al. (2011b), consists of tagging a word as non-relevant or giving it a list of manually defined tags such as lemmas or word classes. For the SMT approach, two methods were evaluated: the Moses system and a Stochastic Finite-State Transducer using the GIATI algorithm (Casacuberta and Vidal, 2004). Both methods use the same alignments between words and signs. However, to improve their Spanish-to-LSE SMT system, Factored Translation Models are combined with the phrase-based SMT model (López-Ludeña et al., 2011a). Factors are semantic categories and linguistic information such as part of speech or morphological features. The best BLEU obtained is above 0.70.

Examining the brief history of machine translation systems for sign language, it seems that research and development of such systems generally followed contemporary machine translation trends. Within rule-based machine translation, emphasis has been placed on the modelling of particular phenomena, using the computational linguistic theories and formalisms available at the time of any given study. Conversely, the data-driven paradigm exploits statistical regularities found in available parallel data, which are scarce for sign languages. Scarceness has led to an attempt to reduce parameters by using semantic classes and by incorporating linguistic information in the form of factors. Doing so has resulted in some level of success, but only in very limited domains. However, parallel data for sign languages are expected to remain scarce in the foreseeable future, unless research efforts are focused on their acquisition.

The following sections contain a description of the different stages of processing from written Spanish to LSE's phonological representations proposed in this dissertation.

3.2 Spanish Analysis

The analysis stage is carried out by free software and linguistic resources available for Spanish. The analyser makes use of the tokenisation and part-of-speech tagging delivered by FreeLing, which also provides the transfer phase with named-entity classification and recognition and word sense disambiguation using EuroWordNet (Vossen, 1998). Unfortu-



(a) Constituency tree

(b) Dependency tree

Figure 3.2: Two kinds of analyses for '*Cuando está solo no se aburre* (When he is alone he does not get bored)'. Syntactic functions are labelled in the constituency tree branches of (a) and on dependency arcs of (b), where arrows go from heads to dependents.

nately, despite FreeLing can provide word sense analysis with a certain accuracy, there are no bilingual lexical resources at the sense level for being used by the system.

Syntactic analysis is carried out by MaltParser (Nivre et al., 2007), a data-driven dependency parser. In dependency parsing, the syntactic structure of a sentence is described in terms of words and binary semantic or syntactic relations between words; constituency or phrase structure does not play any role. Figure 3.2 contains the two kinds of analysis for a sentence. Given a phrase-structure analysis annotated with grammatical functions, it is not difficult to see how it can be transformed into a dependency analysis. The instance of MaltParser used has been trained with the IULA Spanish LSP Treebank, a corpus consisting of almost 590,000 tokens in 42,000 sentences from different technical domains (Marimon et al., 2012). The resulting Spanish model has been evaluated with a 20% held-out set of the IULA Spanish LSP Treebank, which reports a Labelled Attachment Score (LAS) of 93.14% and a Labelled Complete Match (LCM) of 47.60%. LAS is the accuracy obtained in assigning both a head and a function label; LCM is the proportion of completely correct analysed sentences. Results from a newspaper corpus, the Tibidabo Treebank (Marimon, 2010), report an LAS of 88.95% and an LCM of 36.20%. These results represent the state of the art in dependency parsing for Spanish.[3]

3.3 Spanish-Spanish Sign Language Transfer

The transfer stage involves two different sub-stages: lexical transfer and structural transfer. The lexical transfer employs a bilingual LSE-Spanish lexicon to convert the Spanish dependency tree generated from the analysis stage into an intermediate representation between the Spanish and LSE of this dependency tree. The structural transfer converts

[3] <http://www.iula.upf.edu/recurs01_mpars_ul.htm> (Accessed: November 2014)

this intermediate dependency tree into an LSE dependency tree that is then moved to the generation stage. The bilingual lexicon is expanded by the incorporation of morpho-lexical and lexical-semantic relationships.

3.3.1 Lexical Transfer

Resources for translating sign languages into written languages are rather scarce. Fortunately, there are several lexical resources for the translation of written Spanish into LSE glosses. On the Spanish side, there exists a normative dictionary, the DRAE (Real Academia Española, 2001), containing more than 100,000 entries, and analysis lexicons, containing between a half and a million word-forms from between 60,000 to 80,000 lemmas. In the case of FreeLing, the system comes with a Spanish lexicon containing 550,000 word-forms corresponding to 76,000 lemma-PoS combinations. On the LSE side, there is also a bilingual Spanish-LSE electronic dictionary, the DILSE-III (*Diccionario normativo de la LSE*; Fundación CNSE (2008)), which contains LSE videos, Spanish definitions taken from the DRAE, SEA transcripts as well as other linguistic information for about 4,100 combinations of signs and senses corresponding to 2,500 Spanish words. It is the first LSE normative dictionary, created by the CNSE[4]. In addition to the signs of the DILSE-III, all the signs defined in the *Gramática didáctica de la lengua española* (GDLSE) are also considered normative. All these signs, together with their bilingual equivalences in Spanish constitute the basis of the transfer lexicon.[5]

The coverage of this bilingual lexicon as a resource for machine translation is rather low, it contains about 2,400 lemmas. To bridge the gap between language vocabularies, several mechanisms for expanding the bilingual lexicon with new Spanish-LSE entries are proposed in the following sections. All these mechanisms arise from the contrastive study of the two languages: On the one hand, the typological differences between the part-of-speech systems of the two languages can be take into consideration. On the other hand, semantic lexical relations in lexical ontologies can also be used, not only synonymy and hypernymy, but also others for making complex inferences.

Part-of-Speech Flexibility and Morpho-lexical Relationships

Few linguistic studies have been devoted to the topic of 'flexible words', i.e., words that cannot be classified in terms of the traditional lexical categories (verbs, nouns, adjectives or adverbs). Hengeveld, Rijkhoff and Siewierska (2004) distinguish between flexible,

[4] *Confederación Estatal de Personas Sordas* (Spanish Confederation of Deaf People)

[5] HamNoSys transcripts for signs has been obtained from the SEA transcripts in the ES-LSE Apertium Lexicon in `https://svn.code.sf.net/p/apertium/svn/nursery/apertium-es-ssp/apertium-es-ssp.es-ssp.dix` (Accessed: November 2014).

specialised and rigid languages depending on how grammatical functions and lexemes relate to each other. In specialised or rigid languages, each function has a particular lexeme, whereas flexible languages display classes of lexemes that can be used in more than one function without requiring lexical or syntactic derivation. Flexibility generates homography. A flexible word is used for the same communicative functions that other languages employ different 'specialised' words belonging to two or more of the four traditional word classes. These functions are: predicating (verbs), referring (nouns) or modifying (adjectives and adverbs). A major study of flexibility offering a cross-linguistic, descriptive and diverse theoretical approaches can be found in Rijkhoff and van Lier (2013).

Spanish can be considered a specialised language since lexemes or stems bear the meaning of words, while morphemes express grammatical values and determine the part of speech of the word. Spanish has a great productivity in its word formation mechanisms but prefers derivation to other mechanisms as compounding. Basic derivation mechanisms in Spanish are: parasynthesis, prefixation, emotive suffixation, non-emotive suffixation and back formation (Lang, 1990). However, all the different mechanisms suffer from constraints and changes that make all the derivatives of a word not equally admissible.

However, as Herrero Blanco (2009) points out, LSE displays some degree of flexibility in its part-of-speech system. Sometimes the noun, the adjective and the verb are represented by the same sign, as is the case of the sign glossed as DOLOR, which is represented in Spanish '*dolor* (pain)', '*doloroso* (painful)' or '*doler* (to hurt)' depending on its intended function. In LSE, a sign can be as flexible as to be used as a noun, as an adjective or as the verb, without changing its form. Normally, in LSE, it is the semantic class of a sign what determines its flexibility. Abstract nouns, for example, display flexibility in LSE. When a noun names a property or an entity, its form coincides with the adjectival form. In the other cases, when the noun denotes a state, a process or action, or mental phenomenon, its form is shared with adjective and verb. Flexibility of verbs in LSE is not as general as for nouns. To give some examples, the mechanisms for verbal derivation sometimes add a repetition of the complete sign, as in 'VESTIR-SE (dressing one-self)', which derives from 'VESTIR (to dress)'. Other times, the derived sign becomes two-handed, as in 'ADELGAZAR (to slim)', which comes from the adjective 'DELGADO (slim)'. However, most verbs are flexible with nouns and adjectives. Table 3.2 contains the distribution of word classes corresponding to sign glosses found in DILSE-III sharing phonological transcriptions. This table provides some evidence of LSE's flexibility. It is worth noting that A/N class is also a populated class in Spanish, where many adjectives and nouns share the same wordform.

The Spanish words *dolor*, *doloroso* o *doler* have been derived from the same lexical base and are said to belong to the same morphological family. Spanish derivational

Class	Freq.	Example
N/V	57	ESCRITOR (writer) / ESCRIBIR (to write)
A/N	50	FUTURO (future$_A$) / FUTURO (future$_N$)
N/N	16	EMPLEADO (employee) / EMPLEO (employment)
A/V	9	BORRACHO (drunk) / EMBORRACHAR (get drunk)
A/N/V	5	DOLOROSO (painful) / DOLOR (pain) / DOLER (to hurt)
N/N/V	2	RAZONAMIENTO (reasoning) / RAZÓN (reason) / RAZONAR (to reason)
V/V	2	COGER (to take) / RECOGER (to pick up)
A/A	2	REAL (real) / REALISTA (realistic)
A/A/N	1	MEDICINAL (medicinal) / MÉDICO (medical) / MÉDICO (doctor)

Table 3.2: Frequency distribution of word classes for morphologically related signs sharing transcriptions and glossed examples.

processes are carried out mainly by means of affixation. In order to relate sign glosses with their corresponding Spanish lemma derivatives, a finite-state derivational morphology for Spanish has been implemented. These derivational processes in Spanish have been studied in Santana-Suárez, Carreras-Riudavets and Pérez-Aguiar (2004), Santana-Suárez, Carreras-Riudavets and Pérez-Aguiar (2005), Santana-Suárez, Carreras-Riudavets and Pérez-Aguiar (2006) and Lang (1990). The morphology makes an extensive use of regular expressions and rewrite rules of the kind found in Chomsky and Halle (1968), with the form $(\alpha \rightarrow \beta \ / \ \gamma \ \_\_\_ \ \phi)$, where α, β, γ and ϕ are regular expressions. These rules are compiled into finite-state transducers representing regular relations. For example, deverbal nouns can name the action or the effect of the action, like '*comparación* (comparison)', which derives from '*comparar* (to compare)' and is obtained by the application of rewrite rules like (1), which adds the suffix *-ción* to the participial form '*comparada* (compared)':

(1) Deverbal nouns naming actions or effects of actions
Verb[+*part*, +*fem*, +*sing*] ∘ (da → -ción|-ión|-sión|-zón / ___ $)
∘ *spelling-changes* ∘ *Noun*[+*fem*]

The previous rule (1) takes feminine singular participial forms and changes the suffix '*-da*' by some of the suffixes for action or effect nouns. The replacement of strings is performed at the end of the word as expresses the symbol '$', which denotes the end of the word required as a left context condition of the rewrite rule. The transducer in *spelling-changes* carries out the filtering of unaltered strings and the application of morpho-phonological regularisations in the context of the insertion of the new affix. The final composition with feminine nouns discards all other derivative possibilities.

Nouns for names of quality, state and condition are also derived, like '*humanidad* (humanity)' < '*humano* (human)' and denominal adjectives like '*poderoso* (powerful)' < '*poder* (power)', among others.

The derivational morphology is also used to relate Spanish lemmas like '*nacimiento*$_N$

(birth)' with the sign 'NACER (to be born)' through the pre-existing Spanish-LSE bilingual entry $nacer_V$ (to be born) $\leftrightarrow$ NACER (to be born). The most productive and secure derivation rules in Spanish from Real Academia Española (2011, ch. 5) have been implemented, covering: the derivation of nouns of quality or state; eventive nouns; action and effect of action nouns; deverbal agentive, instrumental and locative nouns; derivation of qualifying and relational adjectives; and verbal derivation with specific affixes.

Formally, these morpho-lexical relations are obtained as follows: Let L be the lexicon, the set relating words with their lemmas and parts of speech (word : lemma_{PoS}), let S be the set of sign glosses with their parts of speech (sign_{PoS}) and let D be the regular relation expressed by rewrite rules for morphological derivation. Spanish lemmas and LSE sign glosses are related via the expression:

$$(\pi_2(L) \circ S) \cup_P (L^{-1} \circ (D \cup D^{-1}) \circ S) \tag{3.1}$$

where $\pi_2(L)$ is the right-hand side of the relation L; L^{-1} and D^{-1} are the inverse relation of L and D respectively; and $\cup_P$ is the priority union of regular relations (Roark and Sproat, 2007, ch. 1). In equation *3.1*, the first operand of the prioritised union $(\pi_2(L) \circ S)$ matches Spanish lemmas with LSE sign glosses with the same citation form, producing a relation like:

$$\begin{gathered}\ldots \\ dolor_N : \text{DOLOR}_N \\ comida_N : \text{COMIDA}_N \\ comer_V : \text{COMER}_V \\ \ldots\end{gathered}$$

The second operand $(L^{-1} \circ (D \cup D^{-1}) \circ S)$ relates Spanish lemmas to LSE sign glosses through Spanish derivatives. D relation contains the wordform pairs resulting from the applications of rewrite rules such as (1), with content similar to:

$$\begin{gathered}\ldots \\ dolor_N : doloroso_A \\ \ldots\end{gathered}$$

Joining D with its inverse D^{-1} in $D \cup D^{-1}$ it is just an algebraic way to make the relation in D symmetrical to composition, producing the following content:

$$\begin{gathered}\ldots \\ dolor_N : doloroso_A \\ doloroso_A : dolor_N \\ \ldots\end{gathered}$$

The expression $L^{-1} \circ (D \cup D^{-1}) \circ S$ computes a chain of relations through the composition of the previously presented relations. In the end, some lemmas and signs get connected, those whose citation forms belong to the same word family. This chain of relations is graphically explained with the following example:

$$\frac{\begin{array}{ccc} L^{-1} & D \cup D^{-1} & S \end{array}}{\begin{array}{ccc} doloroso_A : doloroso_A & doloroso_A : dolor_N & \\ doler_V : dolido_V & dolido_V : dolor_N & \end{array} \quad \text{DOLOR}_N}$$

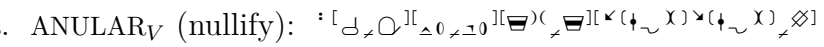
An experiment has been carried out in order to assess to what extent the method proposed here for clustering derivatives is able to reproduce the same classes of phonologically related glosses found in the DILSE-III. To this end, $S \circ (D \cup D^{-1}) \circ S$ has been computed. Let be $\Omega = \{\omega_1, \ldots, \omega_K\}$ the set of clusters induced using the derivative morphology and $C = \{c_1, \ldots, c_J\}$ the set of classes obtained from the lexicon, the quality of the clusters produced is evaluated with the *purity* function, defined as:

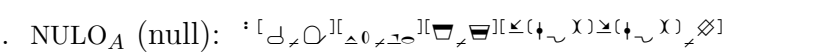
$$\text{purity}(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k \cup c_j| \tag{3.2}$$

Purity values range from 0 to 1, being 0 the worst (impurest) and 1 the best result (purest), i.e., the one with a perfect match between clusters and classes. The purity of the clusters obtained by the method with respect the classes found in the lexicon is 0.703. When impure clusters are analysed individually, very few cases of clusters with etymological unrelated words (*ORAL$_A$ (oral) / ORO$_N$ (gold)) are found since derivation rules implemented have been chosen not to overgenerate. There can be found some clusters with possible but incorrect derivatives, as BORDE$_N$ (edge) / *BORDAR$_V$ (to embroider), which is an interference between derivation, etymology, homography and preferences. Other impure clusters have been found to contain signs with near transcriptions, as the following examples:

(2) a. BAJAR$_V$ (to lower): 

b. BAJADA$_N$ (descent): 

(3) a. ANULAR$_V$ (nullify): 

b. NULO$_A$ (null): 

(4) a. AGRADABLE$_A$ (nice): 

b. AGRADAR$_V$ (to please): 

In example (2) signs differ only in its handshape, corresponding in this case to classifiers (see Table 2.1): a moving person () and a flat surface (). In (3), signs differ in the

dominant hand's initial position (𓂝)(versus 𓂝) and in the movement from this position to the final position, which is identical for both signs. Example (4) differs in the repetition of the movement ($^{+}$) performed in the verbal sign. Reduplication is used to express aspect, number and noun-verb derivation in some sign languages. In American Sign Language (ASL), Supalla and Newport (1978) noted that for a class of semantically related nouns and verbs, verbs are signed with long movements whereas nouns display a restrained and repeated movement. This relation holds for many noun-verb ASL pairs like AIRPLANE/FLY, CHAIR/SIT or IRON_N/IRON_V, concrete nouns often instruments or locations of some activity. However, Schreurs (2006), did not found a consistent pattern in sign language of the Netherlands (NGT) and, in some cases, it was found the opposite pattern to the one found for ASL. In LSE, some common nouns denote collectivities ('PEDREGAL (rocky ground)'), they are signed repeating the classifier. In any case, an in-depth study of derivational relations using the phonological representation of the DILSE-III (also accessible through the LSE-Sign database (Gutiérrez-Sigut et al., 2012)) could be benefit from the clustering of signs using some bounded phonological edit distance.

The bilingual lexicon is increased by over 3,600 new entries relating new Spanish words with previously known LSE signs.

Lexico-semantic Relationships

Another extension of the bilingual lexicon is achieved by the use of lexical semantic relations of the EuroWordNet (EWN) lexical ontology (Vossen, 1998). Apart from the obvious inference from synonymy relations, other inferences can be drawn upon hypernymy-hyponymy relations by allowing for substitution of an unknown word by its hypernym, for example '*coche* (car)' ↔ '*vehículo* (vehicle)'. Let T be the relation with all the pairs of words, with their word class, contained in every synonym set of EWN, and let H be the relation formed by hyponym/hypernym pairs. All these relations are combined to extend the base lexicon L, relating lemmas with no equivalent sign with a semantically related sign, using the following expression:

$$\pi_2(L) \circ \bar{S} \circ (T \cup H) \circ S \tag{3.3}$$

where $\bar{S}$ is the complement of the set of signs S. However, EWN can be exploited further. For example, as (Rodríguez González, 1992) noted, there exist lexical gaps between Spanish and LSE. Some Spanish word, for which there is no equivalent sign, can be validly expressed in LSE by enumerating some of its hyponyms, like '*metal* (metal)' ↔ 'ORO PLATA … (gold silver …)' or '*muebles* (furniture)' ↔ 'MESA SILLA … (table chair …)'. Removing clusters of lemmas with more than five members, the bilingual lexicon can be

extended with 4,200 entries using synonymy relations and over 200 entries relating signs with their hyponyms.

Classifier Nouns

As seen in Section 2.6, a classifier noun is a common noun acting as a free morpheme prepended to another noun to complement its lexical meaning, or to other lexical categories to produce lexical derivatives. The detection of translatable cases using classifier nouns is inferred using a Prolog representation of the Spanish EuroWordNet semantic relationships. For example, for the nouns naming trees, translatable cases are deduced with the hypernym (IS-A) and holonym-part (HAS-A) relationships, which can be formulated in predicate calculus as in (*3.4*), where the translation of '*naranjo* (orange tree)', which is 'ÁRBOL NARANJA (lit. tree orange)', is deduced from the fact that '*naranjo*' is a tree having a part that is the fruit '*naranja* (orange)'.

$$\textit{translation}(x, \text{ÁRBOL } y) \leftarrow \textit{hypernym}(x, \textit{tree}) \wedge \textit{holonym-part}(x, y) \wedge \textit{hypernym}(y, \textit{fruit}) \quad (3.4)$$

For the translation of the names of the inhabitants of a certain place, like '*americano* (American)', the corresponding country name is deduced with (*3.5*), so that it can be translated as 'PERSONA AMÉRICA (lit. person America)'.

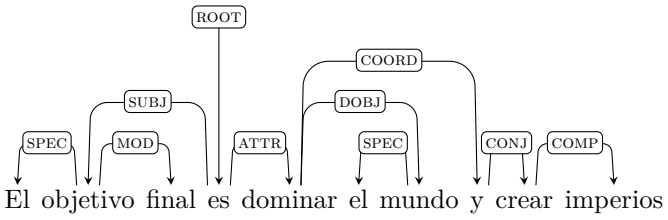
$$\textit{translation}(x, \text{PERSONA } y) \leftarrow \textit{human}(x) \wedge \textit{holonym-member}(x, y) \wedge \textit{derived-from}(x, y) \wedge \textit{place}(y) \quad (3.5)$$

Fingerspelling

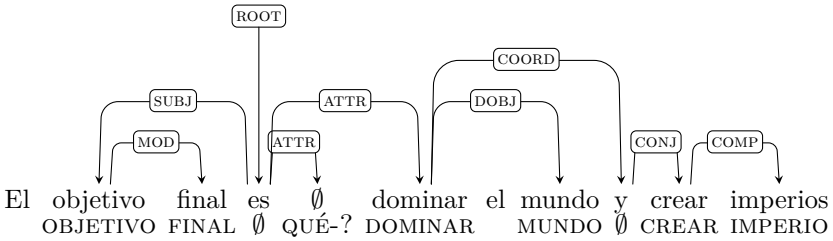
Fingerspelling, which represents alphabets or syllabaries of spoken languages, is used in different languages for different purposes. Fingerspelling or reduced fingerspelling, using only the first letters of the words, may be used as a last resort for representing words from the source language with no equivalent sign in the lexicon, like proper nouns. In so doing, a proper noun like 'Juan (John)' will be glossed 'DL-JUAN' and signed as 'J-U-A-N' or 'J-U-A'.

3.3.2 Structural Transfer

Intermediate dependency trees resulting from the lexical transfer are structurally transferred to LSE. In this process, some nodes representing words are removed, as in the case of definite articles, which have no equivalence in LSE. Other information, however,



(a) Spanish dependency analysis



(b) Intermediate dependency analysis

Figure 3.3: Spanish dependency analysis for '*El objetivo final es dominar el mundo y crear imperios* (The final objective is world domination and empire creation)' and its transferred analysis to LSE where some nodes are disconnected and some new nodes appear. The transferred sentence has been converted into a pseudo-interrogative sentence, which is the proper way to translate a copulative sentence.

can appear in the new structure, like the non-overt subjects in Spanish represented in LSE as personal pronouns whose person and number are inferred from the morphology of the Spanish verb. However, all these transferences are not mandatory, since they are controlled by parameters indicating if the phenomenon has to be transferred. Figure 3.3 contains an example of transference of a copulative sentence into a pseudo-interrogative one in which an interrogative pronoun ('QUÉ (what)') is introduced. Note that definite articles ('*el* (the)') disappear in the transferred tree, since they have no translation as signs in LSE, and that despite the fact that some words are not transferred to signs, their nodes are maintained because their function should be preserved.

3.4 Spanish Sign Language Generation

The generation stage, from an LSE dependency tree, and as a previous step before the synthesis, generates the sequence of glosses. It is divided into two different stages. First, a word order generation takes the LSE dependency tree produced by the transfer stage and produces a sequence of signs representing it; next, a morphological generation takes this sequence of signs and adds the corresponding morphological information. Finally, in order to get the phonological information of a sign all this description is looked up in a

lexicon generated with the computational morphology described in Chapter 4.

3.4.1 Word Order Generation

Another consequence of the Spanish inflectional system is that its constituent order is freer than in other languages with impoverished morphologies, like English. LSE order, as has been seen in Section 2.8, is determined by the verb type and the topicalisation of complements. It is said that word order in LSE, from a functional point of view, is freer than in Spanish, but at the same time, informative order is stricter. Expression of place and time at discourse level differs in both languages, being less redundant in LSE than in Spanish. LSE places this frame information at the beginning of the discourse and does not repeat it in subsequent sentences. On the other hand, Spanish carries temporal/tense information in each verb. In the case of locatives, LSE situates this information in the linguistic space and uses it for agreement in subsequent references. LSE can also use the passive hand as a location for the object signed with the active hand. According to the GDLSE, the linear precedence for constituents in LSE predicative sentences is:

(5) Unmarked word order in LSE sentences
Tc $\prec$ To $\prec$ S $\prec$ IO $\prec$ DO $\prec$ V $\prec$ T/L $\prec$ Asp $\prec$ Mod

where:

- Tc is the 'contextual theme', consisting of generic temporals and locatives,
- To is the topicalised argument,
- S is the subject,
- IO is the indirect object,
- DO is the direct object,
- V is the verb,
- T/L are 'specific' temporals and locatives and
- Asp and Mod are aspect and mood auxiliaries.

The difference between contextual and specific temporals/locatives depends on their roles in the argument structure of the verb. If they are complements, they usually appear after the verb, otherwise they constitute the frame of the predication. LSE has no copulative verbs (linking verbs) and hence copula is expressed as a pause between the subject and the attribute, accompanying the attribute with an assertive movement of the head and

frequently doubling the subject with a pronoun, as in 'JUAN MÉDICO ÉL (lit. Juan doctor he)'.

Broadly speaking, in a predication, the position of its head, normally a verb, is taken as a reference for ordering the rest of the constituents according to their function in the predication. Non-argumental temporal and locative adjuncts appearing before the verb are fronted to occupy the contextual thematic positions. Topicalised or dislocated arguments are heuristically detected as those appearing in front of the verb. They occupy the second position in the LSE sentence, after the contextual theme and in front of the subject and the objects of the verb. These arguments are advanced from positions after the verb to prior positions. All other prepositional complements and adjuncts after the verb are left behind the verb. Modal and aspectual markers are moved to their appropriate positions. SVO is a marked order in LSE that is found in verbal predications with effected or human objects.

Noun phrases constitute another of the major structures in LSE and their constituents have the following linear precedence:

(6) Unmarked word order in LSE NPs
NCL $\prec$ N $\prec$ Compl $\prec$ Dem $\prec$ Poss $\prec$ Num $\prec$ Loc $\prec$ CL-compl $\prec$ Indef

where:

- NCL are classifier nouns,
- N is the head of the NP,
- Compl are the nominal complements and adjuncts of N,
- Dem are demonstratives,
- Poss are possessives,
- Num are numerals,
- Loc are locatives,
- CL-compl are complements with classifiers and
- Indef are indefinites.

However, in the case of marked structures, the basic order differs from the previously established order. For example, 'Dem N' order is used for temporal and locative nouns in the contextual theme, 'Num N' is used for units of measurement and currency names, and 'Poss N' in nouns naming permanent relationships. It is also considered the Spanish

noun phrase with the structure 'N_1 *de* N_2 (N_1 of N_2)'. When translating this expression to LSE, sometimes N_1 precedes N_2 ($N_1 \prec N_2$), as in '*hermano de Luis* (Luis's brother)', which is translated as 'LUIS HERMANO (lit. Luis brother)' because '*hermano* (brother)' denotes a permanent relationship. In other cases, the order is 'N_2 N_1', as in '*mesa de estudio* (study table)', because in this case '*estudio* (study)' is the function or purpose of '*mesa* (table)' and it is thus translated as 'MESA ESTUDIO (lit. table study)'. Meronymy relations, i.e., part-whole relations, usually expressed as 'N_1 de N_2', are translated as 'N_2 N_1', as in '*el volante del coche* (car's steering wheel)', translated as 'COCHE VOLANTE (lit. car steering wheel)'.

Since Spanish and LSE are typological distant languages, compound and complex sentences differ in their surface structures. The GDLSE devotes an entire chapter to the contrast of these structures in both languages. In Spanish, coordination is realised by means of conjunctions, but the LSE's corresponding structure is the juxtaposition of coordinated members with some non-manual components (NMCs) added, as in (7):

(7) a. *Ayer yo trabajé y fuí al cine*
'Yesterday, I worked and went to the cinema'

b. AYER PRO.1 TRABAJAR / $\overline{\text{CINE IR}}^{\text{coord}}$
yesterday I work / cinema go

In the case of Spanish coordinating conjunctions *también*, *tampoco*, *ni* are not translated as conjunctions but some signs are added at the end, like in the following examples:

(8) a. *Yo sé inglés y Luis también*
'I know English and Luis too'

b. PRO.1 INGLÉS SABER / LUIS TAMBIÉN
I English to-know / Luis too

(9) a. *No sé inglés y Luis tampoco*
'I don't know English and neither do Luis'

b. PRO.1 INGLÉS SABER-NO / LUIS TAMPOCO
I English to-know-not / Luis neither

Temporal coordination expressed with '*y* (and)' and with '*y después* (and later)' is translated using the nexus sign DESPUÉS, like in (10).

(10) a. *Voy a dejar el coche en al taller y después iré a casa*
'I'm going to leave the car at the garage and then I'll go home'

b. PRO.1 COCHE TALLER DEJAR / DESPUÉS CASA IR
I car garage left / later home go

However, conjunctive expressions '*pero antes* (but before)' marking previous actions are translated introducing the sign 'PRIMERO (first)' in the same position as 'DESPUÉS'.

Spanish counterfactual disjunctions are translated into a kind of conditional with negative effects:

(11) a. *Debes ir a comer o no irás a pasear*
'You must go to dinner or you won't go to walk'

b. PRO.2 COMER DEBER / EJEMPLO NO / PASEAR NO
you to-dinner must / example no / to-walk no

Causal, final and consecutive sentences are translated into logical coordination. In the case of causal sentences, the signs 'MOTIVO (reason)', 'PORQUÉ (because)' or 'POR (because of)' are used like in the following sentence:

(12) a. *Me he comprado un coche nuevo porque el viejo está roto*
'I have bought a new car because the old one is broken'

b. PRO.1 COCHE NUEVO COMPRAR / MOTIVO / COCHE VIEJO ROTO
I car new to-buy / reason / car old broken

Final nexus in LSE are 'FINALIDAD (end)' or 'OBJETIVO (objective)', with the same pauses and intonation as for causals:

(13) a. *Luis se ha apuntado a clases particulares para aprender inglés*
'Luis has signed up for private classes to learn English'

b. LUIS CLASE PARTICULAR APUNTAR / OBJETIVO / INGLÉS APRENDER
Luis class private sign-up / objective / English to-learn

In Spanish, consecutive conjunctions are '*si* (if)', '*entonces* (then)', '*en consecuencia* (consequently)', '*por eso* (therefore)' and others. They are realised by means of an NMCs emblem represented by '[ASÍ-QUE] (so)':

(14) a. *No sé inglés, así que encontrar trabajo es difícil*
'I don't know English, so to find a job is difficult'

b. PRO.1 INGLÉS SABER-NO / [ASÍ-QUE] / TRABAJO ENCONTRAR DIFÍCIL
I English to-know-not / so / work to-find difficult

Relative clauses in Spanish are subordinate clauses introduced by the pronouns '*que* (what)', '*quien* (who)', the adverb '*donde* (where)' and others. Translating this kind of subordinated sentences into LSE involves:

- the topicalisation of the relativised object within the relative clause
- the topicalisation of the relative clause

- the insertion of a pause between the sentential topic and the dependent clause
- an ascending intonation

This double topicalisation is shown in example (15). Note that the simple sentence '*Ayer vino un hombre* (A man came yesterday)' is translated as 'AYER HOMBRE VENIR (lit. yesterday man to-come)'.

(15) a. *Le he dado un libro al hombre que vino ayer*
'I've given a book to the man that came yesterday'

b. HOMBRE AYER VENIR / PRO.1 LIBRO DAR
man yesterday to-come / I book give

Another kind of subordinate clauses is clausal complements. They are realised in the same place as in Spanish without the complementiser or as pseudo-interrogatives. Both forms of realisations are exemplified respectively in (16) and (17):

(16) a. *Luis ha visto que Juan llora*
'Luis has seen that Juan cries'

b. LUIS VER / JUAN LLORAR
Luis to-see / Juan to-cry

(17) a. *Luis ha visto que Juan llora*
'Luis has seen that Juan cries'

b. LUIS VER $\overline{\text{QUÉ}}^{?}$ / JUAN LLORAR
Luis to-see what-? / Juan to-cry

Finally, hypotheticals in LSE are realised by introducing the sign 'EJEMPLO (example)' or with CNMs. Spanish conditional, concessive, article + relative and prepended final constructions are translated as conditionals. The following are examples of the aforementioned constructions:

(18) a. *Si mañana llueve, no iré a la playa*
'If it rains tomorrow, I won't go to the beach'

b. EJEMPLO / MAÑANA LLOVER / PRO.1 PLAYA IR NO
example / tomorrow to-rain / I beach to-go not

(19) a. *Aunque llueva, iré a trabajar*
'Although it rains, I'll go to work'

b. EJEMPLO LLOVER / DA-IGUAL / PRO.1 TRABAJO IR
example to-rain / it-does-not-matter / I work to-go

(20) a. *El que llegue tarde debe avisarme*
'The one who arrives late have to warn me'

b. EJEMPLO / LLEGAR TARDE / AVISAR-CONC.1 DEBER
example / to-arrive late / warn-me have-to

(21) a. *Para conseguir el premio, tendrás que sacrificarte*
'To get the prize, you will need to sacrifice yourself'

b. EJEMPLO / PREMIO CONSEGUIR / PRO.2 SACRIFICAR
example / prize to-get / you to-sacrifice

The ordering of signs is formulated as a constraint satisfaction problem. A precedence graph is built from the information about the relative order between a head and its dependent or between pairs of dependents sharing a head. In (22), an example Prolog rule is shown stating that the prepositional modifier of a noun precedes it if the head noun holds a holonomy-meronomy with the object (COMP) of the prepositional phrase.

(22) noun(holonym) > PP[de, noun(meronym)]

```
lp(H, _, D, >) :-
  linear_precedence(lse, std),
  tok(H, lemma, Head),
  tok(D, lemma, 'de'),
  tok(O, head, D),
  tok(O, deprel, 'COMP'),
  tok(O, lemma, Obj),
  part_whole(Obj, Head), !.
```

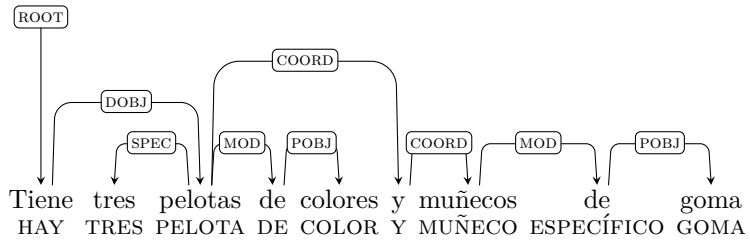
For the generation of constraints on word order from the dependency tree, the algorithm applies the following steps:

1. For each word H, we create two convenience elements '$(_H$' and '$_H)$', which represent the boundaries of the phrase headed by H, which maintains the relation '$(_H \prec H \prec {}_H)$'.

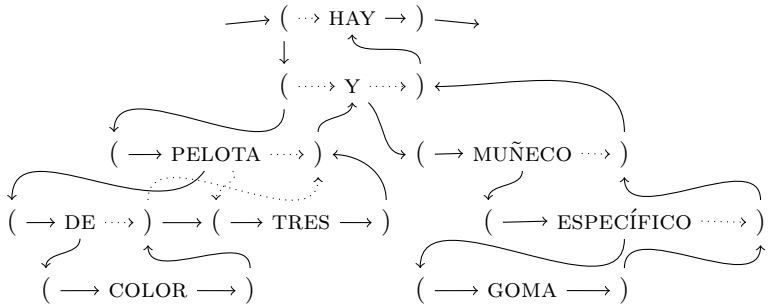
2. If a dependent D of a head H is topicalised, '$(_H \prec (_D$' and '$_D) \prec H$'.

3. For every pair of dependents D_1 and D_2 of a head H, if D_1 is topicalised, then '$_{D_1}) \prec (_{D_2}$'.

4. For every non-topicalised dependent D of a head H, if D precedes H then '$(_H \prec (_D$' and '$_D) \prec H$', and if H precedes D, then '$H \prec (_D$' and '$_D) \prec {}_H)$', otherwise '$(_H \prec (_D$' and '$_D) \prec {}_H)$'.



(a) Dependency analysis



(b) Precedence graph

Figure 3.4: Dependency analysis of '*Tiene tres pelotas de colores y muñecos de goma* (He has three coloured balls and rubber dolls)' and precedence graph from its output from the transfer stage. The circuit outlined with solid lines represents a topological ordering of the graph that will result in the gloss sequence 'PELOTA DE COLOR TRES Y MUÑECO ESPECÍFICO GOMA HAY (lit. ball of colour three and dolls specific rubber there-is)'.

5. For every pair of non-topicalised dependents D_1 and D_2 of a head H, which has been placed on the same side in the previous step, if D_1 precedes D_2, then '$_{D_1}) \prec (_{D_2}$' and if D_2 precedes D_1, then '$_{D_2}) \prec (_{D_1}$'. Otherwise, they maintain the same relative order as in the source sentence.

6. Finally, apply topological sort to the graph of precedence, breaking any detected cycles, obtaining the linear ordering of signs.

Since step 6 of the previous algorithm can only be applied on an acyclic graph, errors during development on inconsistencies are easily detected as cycles and reported to the grammar writer. Figure 3.4 contains an example of the application of the algorithm to a dependency analysis. As can be seen in Figure 3.3a, the dependency structure for the transferred tree is identical to the original tree but not to the transferred lexical items. The gloss sequence is obtained by applying a topological ordering to the precedence graph in Figure 3.4b, generated with the previous algorithm. It is worth noting that conditions in step 4 of the algorithm ensure non-projectivity in the constituency of the realised sentence.

3.4.2 Morphological Generation

Some glosses not realised as signs are simply removed, but some other glosses are annotated with morphological features that modify their phonological representation, like an absent third person personal pronoun located to the left of the signer, a transitive verb agreeing with its object, a plural realised by repeating the noun, or a verb with a continuous aspect. However, this issue will be addressed in length in Chapter 4.

Special cases in generation are classifier predicates. Some LSE postpositions are mainly nominal complements with locative or temporal meanings. In Spanish, these complements correspond to nominal phrases with prepositional phrase modifiers introduced by locative prepositions and locutions as '*sobre* (on)', '*bajo* (under)' or '*tras* (behind)' and by temporal prepositions as '*entre* (between)' or '*desde* (from)'. In these cases, the locative construction can be generated as a two-handed sign in which the passive hand represents the head of the phrase and the active hand represents the term of the preposition, both with their entity classifier.

3.5 Evaluation

Human evaluation is fundamental and remains crucial for the proper assessment of the quality of machine translation systems. When a sign language animation from the output of the machine translation system is evaluated, the whole process is taken into account – not only the translation into a symbolic representation, but also its synthesis as an animation, how realistic and fluid the performance of the avatar is, etc. The fact remains, however, that the lack of an accepted writing system for sign languages poses problems to machine translation evaluation. Despite the existence of transcription systems, much vital information for the production of signed utterances is missing in any text-based representation and automatic evaluation using distances between the reference and candidate translations as mere strings is inadequate for measuring how much of the essence of the reference translation is preserved (Morrissey and Way, 2006). In addition, the choice of the transcription representation influences the evaluation result. SiGML (Signing Gesture Markup Language), glosses and sign identifiers were used by Morrissey (2011) to explore this effect. Although the scores achieved by SiGML were found to be better than those achieved by glosses, and despite the fact that glosses are not considered an adequate representation of signs by many, automatic evaluation of text-based representations remains useful to assess the internal progress of machine translation systems. In addition, the results obtained could be comparable with those found in the research literature and can lead to a correct sign language output with a synthesis module using a lexicon relating glosses and animated forms.

3.5.1 LSE Corpus Description

A parallel Spanish-LSE corpus has been created from the material used in the psycholinguistic study published in Rodríguez Ortiz (2005). The study seeks to assess the effectiveness of LSE as a language of instruction compared with an oral language and the effect of late learning of LSE on comprehension. An experienced human interpreter was in charge of translating the texts originally written in Spanish into LSE. The performance was first video-recorded and then transcribed into a gloss notation. A bimodal bilingual second interpreter (daughter of two deaf parents) translated the video recordings in LSE back into Spanish. The resulting Spanish speech was then manually transcribed. We will call the original Spanish texts corpus A; the LSE glosses resulting from the transcribed video performance corpus B; and the transcription of the speech translation of the video-recorded performance corpus C. According to the author of the study, the meaning of the back-translated sentences in C agrees quite well with their correspondent sentences in A, confirming both the skill of the first interpreter in translating into LSE and the naturalness and fluency of the Spanish in the back-translation performed by the second interpreter.

Experiments were conducted using six different texts with different topics and with different degrees of difficulty in understanding. The topics covered by texts one to six are as follows: student accommodation, history and war, a day in a dog's life, population and demographic problems, domestic violence, and a day in a little girl's life. The difficulty of a text is determined by the background knowledge required to understand it, the topic variation, its thematic progression, the structure of each text, the ease of extraction of the main ideas, the domain vocabulary, the lexical richness and the degree of abstraction. Texts one and two are of low difficulty, texts three and four are of average difficulty and texts five and six are the most difficult to understand.

All the texts of the corpus come in three versions: (A) the Spanish source version, (B) an LSE translation of A and (C) a Spanish translation of B. The three versions of the corpus have been manually segmented into sentences and aligned appropriately. As an example, (23) contains three aligned sentences extracted from the corpus. The sentence in (23a) is the source sentence in A, the sentence in (23b) corresponds to the LSE translation of (23a) in B and that, in turn, is translated back into Spanish in (23c) in C.

(23) a. *La guerra es tan vieja como la humanidad*
'War is as old as humanity'

b. GUERRA EDAD IGUAL HUMANIDAD IGUAL+
war age equal humanity equal-ASP.CONT

c. *La guerra tiene la misma edad que la humanidad*
'War has the same age as humanity'

			Words/Signs			Punctuation	
	Language	Sentences	Tokens	Types	Lemmas	Tokens	Types
Corpus							
A	Spanish	229	3,033	1,082	782	373	9
C	Spanish	229	2,946	912	642	446	8
B	LSE	229	1,992	611	-	348	2
Testbed							
A	Spanish	195	2,538	957	700	307	7
C	Spanish	195	2,519	813	580	394	8
B	LSE	195	1,709	532	-	-	-

Table 3.3: Counts of sentences, tokens (or running items), types (or vocabulary), lemmas and punctuation signs (excluding full stops) for versions A, B and C of the texts in Rodríguez (2005).

The number of sentences and lexical items is shown in Table 3.3. In Spanish counts, for versions A and C, word-forms, multiwords, numbers, quantities, punctuation, dates, proper names, etc. are included. LSE counts for version B include signs, classifier predicates as a unit and punctuation. The average number of tokens per sentence in Spanish is fourteen to fifteen, while the average number of tokens in the LSE sentences is approximately ten.

The LSE part of the corpus has been annotated following the guidelines published in Alonso Baixeras et al. (1998). Glosses are accompanied by several symbols describing prosodic, syntactic and morphological values. Classifier predicates are also tagged and described. Unfortunately, marking has not been applied uniformly and consistently. Not all the instances of the same phenomenon have been marked and not all the marked instances have always been marked in the same way. The symbol '(?)' marks pseudo-interrogatives, which are realised by means of NMCs. The symbol '(cond)' is a mark used in conditionals, also performed by NMCs. The single or multiple repetitions of a sign are marked using the symbols '+', '++' and '+++'. However, they are not used consistently, even within the same text. The most frequent use of the symbol '+' corresponds to the plural marking in nouns, like 'PERSONA+ (people)', to some kind of adverbial intensification in verbs like 'PREOCUPAR+ (to worry-INTENS)', or to the aspectual marking in verbs, like 'BUSCAR+ (to search-ASP.FREQ)'. The symbol '++' is also used for quantifying in either nouns like 'PERSONA++ (people)', adjectives like 'TRAVIESO++ (naughty-PL)' or verbs like 'JUGAR++ (to play-ASP.CONT)'. The symbol '+++' seems to be used to mark descriptive plural, i.e., a kind of plural realised by repeating the signs in different places, sometimes in locations analogous to the locations of these objects in the real world. The symbol '(mofl)' stands for an NMC realised by inflating the cheeks with a value of intensification. Finally, the symbol '(2m)' states that one-handed signs

have been made two-handed, as a realisation of plurals, aspectual values, or reciprocity, among other possibilities. There are different marks for some kinds of classifier predicates: descriptive constructions (CLD), locative constructions (CLL), pronominal constructions (CLS), body-as-subject related actions (CLC), body-part related actions (CLCP) and instrumental constructions (CLI). The corpus contains thirty-four classifier constructions, but no one contains the information for its articulation.

3.5.2 Experiments

To evaluate the system output quality, a testbed has been created using the corpus described in the previous section. Because Spanish sentences in C are obtained from LSE sentences in B and B sentences are obtained from Spanish sentences in A, C and A can be considered paraphrases, since C is a back-translation of A. However, from lexical and structural points of view, the differences between A and B and the differences between B and C are not the same. It is worth noting that some kind of interpretation has been applied in the translation from A to B. The existence of classifier predicates in the LSE translation supports that hypothesis. Both versions A and C convey the same meaning, however, the translation from B to C is a more 'literal' translation in lexical and structural terms and B seems *a priori* easier to obtain from C than from A. In order to avoid the introduction of artificial differences, morphological mark-up is not taken into account, since there is not enough consistency in its use. Sentences containing classifier predicates have not been incorporated into the testbed, although the possibility of generation of these classifier predicates will be discussed later. Each testbed used for experimentation consists of 195 sentences. Counts for words and signs can be found in Table 3.3.

In automatic machine translation evaluation, the most widespread metric used is BLEU (Papinemi et al., 2002). BLEU is calculated by combining individual n-gram precision of up to four with a brevity penalty. Translation error rate (TER) is another common metric used for machine translation evaluation. TER attempts to measure the minimum amount of editing that a human would have to perform to convert the system output into the reference translation (Snover et al., 2006). The metric is computed from the number of insertions, deletions and substitutions of words, and from the number of phrasal shifts, i.e., the movement of word sequences to other locations of the output translation. Note that while BLEU is better if higher, TER is better if lower.

In order to evaluate and compare the RBMT approach presented in this thesis with other approaches, several experiments have been conducted with the phrase-based Moses translation model (Koehn et al., 2007) providing a baseline with a data-driven approach. For Moses, initial experiments with development sets led to poor results because of the scarcity of the data of the testbed and the sparseness of the problem, which cause over-

fitting. These poor results agree with those reported by Morrissey (2011) as having been obtained in similar experiments. Results for Moses are estimated using cross validation.

In the experiment called Experiment I in Table 3.4, standard ten-fold cross-validation is used, splitting data randomly into training and test. Both the language model and the phrase translation table are obtained from the training set of each fold with GIZA++ (Och and Ney, 2003). As it can be seen from the results in Table 3.4 BLEU is zero. The reason for this low result is that three- and four-gram precisions are zero for almost all the test folds. An according high average TER indicates that much postediting work on the output will be required to match the reference translation. Moreover, a TER higher than 100% indicates that there are more words in the output translation than in the corresponding reference translation. When inspecting the phrase table obtained, it seems clear that, due to the experiment's small size, data are incorrectly aligned, causing most of the errors. This result applies to both testbeds.

In Experiment II, six-fold cross-validation is used but in each fold one of the six texts is left out the training set and used as test set. The second experiment aims to evaluate how well Moses works when training and test domains differ when translating Spanish into LSE. As can be seen in Table 3.4 for both testbeds, modified n-gram precisions are slightly lower and TER is slightly higher than in Experiment I. This result might indicate some domain dependency but a reliable conclusion should be based on more data.

For the third and fourth experiments using Moses (Exps. III and IV in Table 3.4), different language models are obtained from the training data of each fold, as in the previous experiments. However, in an attempt to improve the results obtained with GIZA++ in Experiments I and II, Moses is provided with a phrase table obtained from the bilingual lexicon of the RBMT system, i.e., the bilingual lexicon. In that table, the same probability has been given to every possible translation of a word. The phrase table is the same for each of the folds. From the results, the good average adequacy obtained from the cross-validation tests in light of the precision figures for shorter n-grams should be noted. The use of good and accurate bilingual resources, like bilingual lexicons or manually aligned data, is clearly helpful to SMT when parallel text data are scarce. Similar results are obtained within both testbeds.

Experiment V uses the RBMT system of this thesis without the generation module described in Section 3.4.1. In this case, the generation algorithm has been replaced by an algorithm preserving the original order of the corresponding words in the Spanish source sentence. The resulting order looks like Signed Exact Spanish (SES), which is not a natural language but a manually coded language, i.e., a representation of an oral language in a gestural-visual form. However, the output differs from SES mainly in that, because of the lexical transfer, not all Spanish words are expressed in LSE and the correspondence

Exp.	Testbed	TER	BLEU	Modified N-gram Precision				Methods
				1-gram	2-gram	3-gram	4-gram	
I	A	140.31	0.000	0.152	0.012	0.002	0.000	Moses + GIZA++ + crossval
II	A	144.14	0.003	0.142	0.009	0.002	0.001	Moses + GIZA++ + crossdom
III	A	82.45	0.018	0.394	0.082	0.021	0.004	Moses + Bilingual Resources + crossval
IV	A	82.85	0.033	0.384	0.086	0.022	0.004	Moses + Bilingual Resources + crossdom
V	A	72.56	0.088	0.492	0.157	0.058	0.019	RBMT without parsing + Span. word order
VI	A	72.21	0.089	0.492	0.164	0.058	0.019	RBMT + MaltParser

Exp.	Testbed	TER	BLEU	Modified N-gram Precision				Methods
				1-gram	2-gram	3-gram	4-gram	
I	C	116.06	0.000	0.260	0.026	0.003	0.000	Moses + GIZA++ + crossval
II	C	121.03	0.005	0.237	0.017	0.003	0.002	Moses + GIZA++ + crossdom
III	C	46.47	0.217	0.765	0.332	0.157	0.082	Moses + Bilingual Resources + crossval
IV	C	46.90	0.216	0.762	0.325	0.154	0.082	Moses + Bilingual Resources + crossdom
V	C	41.61	0.296	0.801	0.405	0.244	0.154	RBMT without parsing + Span. word order
VI	C	42.73	0.302	0.800	0.410	0.252	0.159	RBMT + MaltParser
VII	C	41.96	0.306	0.801	0.417	0.256	0.162	RBMT + MaltParser + Manual Corrections

Table 3.4: Results of the experiments: BLEU and precision values are given as indices but the TER metric is expressed as percentage. A TER higher than 100% indicates that there are more words in the output translation than in the corresponding reference translation. Note that while BLEU value is better if higher, TER is better if lower.

between words and signs is not one-to-one. It is worth noting that similar experiments can be found in the literature. In Kanis et al. (2007), using a training set of 12,616 sentences, translation from Czech to Signed Czech, a manually coded language, reached a BLEU of 0.81, a WER (Word Error Rate) of 13.14% and a PER (Position Independent Word Error Rate) of 11.64%. In Stein, Schmidt and Ney (2010), two experiments applied to both German and DGS involving simple lowercasing and four-letter stemming have been conducted. The BLEU/PER obtained in each experiment was 0.021/85.7% and 0.026/81.1% respectively. However, the reported baseline with Moses was 0.181 BLEU and a 71.0% TER with a training set of 2,565 sentences and a test set of 512 sentences. Combining several systems, they finally reached a BLEU of 0.234 and a TER of 65.5%. Note that the disparity between these results is because Czech and Signed Czech have the same surface order, but German and DGS do not. Here, figures for the RBMT system are given in Table 3.4. However, the real reason is that, as explained in Section 3.5.1, source sentences in the testbed C have been obtained as back-translations from LSE sentences and a slight bias towards LSE surface order can be detected, which is perfectly accommodated in a relatively free word-order language like Spanish. This could explain the relatively good BLEU obtained in this experiment.

The complete RBMT system from Spanish to LSE glosses described in this thesis has been evaluated as Experiment VI in Table 3.4. This system has been provided with several parameters to deal with the specific LSE variation found in the corpus. These

parameters are used in the experiment to approximate linguistically the system output to its corresponding reference translation in the testbed. In view of the results, there is a general gain in terms of BLEU and TER with respect to the figures obtained in the rest of the experiments. However, the most important thing to note here is that, in order to get better results than in Experiments III and IV, many implemented word order rules were deactivated. These rules were implemented according to the LSE grammar but the order of signs of the testbed does not seem to follow the current norm in LSE.

In order to quantify the effect of the parser quality on results in terms of BLEU and TER, many dependency trees delivered by the parser have been manually corrected. The number of corrections applied to testbed C totals 399 distributed as follows: 220 head assignments and 179 function labels. Corrections affected 263 tokens (a 7.7% of the tokens). The results of this experiment can be seen in Experiment VII. The improvement in BLEU and TER with respect to the uncorrected parses is very small.

Results confirm that data scarcity and domain sparseness lead the data-based approach to perform worse than the rule-based system. Providing bilingual lexical resources have a positive effect in data-based approaches but differences using cross validation or leaving one text out are small. We think that this result should not be interpreted as domain independence. Instead, we consider that data are not still enough to measure the out-of-domain effect. On the rule-based translation side, the most important conclusion that can be drawn from the experiments is that the order of signs is similar to the order of words in the Spanish fragments in C. We think that this result should not mean that LSE and Spanish have similar word orders or that the order generated by the system is not valid. Instead, this result supports the hypothesis that the word order found in Spanish testbed C follows the order of LSE sentences in B, since the correction of the output of the parser does not have a dramatic improvement in translation results. At this point, deeper and more extensive experiments measuring human understanding should be performed to draw reliable conclusions.

3.5.3 Analysis of Errors

Because data-driven approaches base much of their success on data, an analysis of errors is rarely found in evaluations. However, a proper classification of errors, which should take into account the linguistic phenomena involved in the error, can be helpful in choosing an appropriate system architecture and in the development of new rules for RBMT systems and resources for translation. As an exception, López-Ludeña et al. (2012) contains an analysis of errors for the Spanish-to-LSE SMT system described in Section 3.1 applied to a highly specific domain. In that work, the main sources of errors are reported in order of decreasing importance as follows: (a) the differences in the number of words and

signs for parallel sentences, due to the absence in Spanish of pronominal subjects and the absence in LSE of definite articles, prepositions and copula, and the different realisations of plural, and others; (b) the differences in word order (SVO versus SOV); (c) the incorrect generation of classifier predicates; (d) the out of vocabulary words; and (e) the specific names in LSE periphrastic expressions.

Apart from the expected differences caused by the lexical gap, a detailed analysis of errors between Experiment VI and VII reveals that some differences are caused by the incorrect assignment of functions in the dependency analysis. However, many other differences come from the choice of alternative structures to convey meaning. Some examples are reported below, showing the source sentence 'a.', the target or reference translation 'b.' and the system's output 'c.'. Sometimes the differences arise from an alternative form of realisation, as in (24), where the intensification of a property as in '*muy feliz* (very happy)', which plays the role of a predicative complement in '*lo vemos muy feliz* (we see him/her very happy)', is translated by doubling the predicate, resulting in 'VER FELIZ / VER FELIZ (lit. see happy, see happy)', which is a more iconic expression of quantity.

(24) a. *Cuando abrimos lo vemos muy feliz*
'When we open, we see him/her very happy'

b. TOCA ABRIR / VER FELIZ / VER FELIZ
turn to-open / see happy / see happy

c. TOCA ABRIR FELIZ MUY VER
turn to-open happy very see

Other examples in the testbed also display variation in the choice of referring expressions. For example, the Spanish female reference '*Ana* (Ann)' is realised either as the finger-spelled sign 'DL-ANA (A-n-n)', or as the deictic 'PRO.3 MUJER (she woman)', a pronoun with a free morpheme carrying the information about the female sex of the referent, or as the generic 'MUJER (woman)'. All these three forms of realisation can be seen respectively in examples (25)–(27).

(25) a. *Allí Ana juega con sus amigos*
'Ann plays there with her friends'

b. ALLÍ PARQUE AMIGO DL-ANA JUGAR
there park friend A-n-n to-play

c. ALLÍ ANA JUGAR AMIGO SU
there A-n-n to-play friend her

(26) a. *A Ana le encanta la bata*
'Ann likes her dressing gown'

b. PRO.3 MUJER ENCANTAR BATA
he woman to-like dressing gown

c. BATA ANA ENCANTAR
dressing gown Ann to-like

(27) a. *Ana no hace nada*
'Ana does nothing'

b. MUJER HACER NADA
woman to-do nothing

c. ANA NADA NO-HACER
Ann nothing not-to-do

It is worth noting that example (27) contains a double negative, often required in Spanish, whose translation in (27c) does not match the reference in (27b). However, as in other languages, the double negative could be understood as a form for adding emphasis to the negative interpretation. In any case, the translation of the general negation is not fully implemented, because of the complex interactions of negation with other phenomena, which require further study.

Complex sentences, i.e., those with an independent clause and one or more dependent clauses, are stable in Spanish. However, according to GDLSE, LSE presents alternative forms of expression containing non-standard markers, which are still in the process of grammaticalisation and hence have an uncertain future. This issue causes differences as seen in (28), where the clause structure is marked in square brackets. The first dependent clause is introduced by the complementiser '*que* (that)', which plays the role of direct object of the verb '*querer* (to want)'. The second clause, a causal clause, is introduced by '*porque* (because)'. The output translation in (28c) reflects the original structure in (28a), but differs significantly from the structure of the reference translation in (28b), in which both the causal marker and the verb are doubled.

(28) a. [*Mamá no quiere* [*que los niños vayamos a la calle*] [*porque está oscuro*]]
'Mom does not want children to go to the street because it is dark'

b. [MOTIVO MAMÁ QUERER-NO [NIÑOS IR CALLE] QUERER-NO] [MOTIVO OSCURO]
[reason mom not-to-want [children to-go street] not-to-want] [reason dark]

c. MAMÁ [NIÑOS IR CALLE] QUERER-NO [MOTIVO OSCURO]
mom [children to-go street] not-to-want [reason dark]

Other differences arise from the interaction of several phenomena. A very common construction found in sign languages is rhetorical questions or pseudo-clefts. Structurally, they consist of a question-answer pair and are used to bring a particular constituent into

focus. An example of this construction can be found in (29b). In the example, several markers are introduced. First, '-?' is used to represent the NMCs expressing a question and then '/' represents an intonation break. Note that, in this particular example, because of the interaction of a coordinating conjunction, the structure appears twice.

(29) a. *El objetivo final es dominar el mundo y crear imperios.*
'The final objective is to dominate the world and to create empires'

b. OBJETIVO FINAL QUÉ-? / MUNDO DOMINAR / CREAR QUÉ-? / IMPERIO
objective final what?, world to-dominate, to-create what empire

c. OBJETIVO FINAL MUNDO DOMINAR IMPERIO CREAR
objective final world to-dominate empire to-create

There are several possible reasons for some of these differences. First, a single non-native signer has done the translation and, therefore, several aspects could have influenced the native language of the signer, such as word order. Second, data were acquired in Seville, in southern Spain, away from the influence centres of the standard, which are thought to be central and eastern communities. Third, when the corpus was created in 1995, no standard existed for LSE. Therefore, the language used surely corresponds to a variety with notable differences from the current LSE standard. And fourth, some degree of variation exists within LSE, as in all languages, which leads to the existence of different alternative forms of realisation of the same linguistic phenomenon. Being those different forms equally valid, some parameters could be introduced to the system in order to select one form of realisation or another.

3.5.4 Analysis of Classifier Predicates

Classifier predicates deserve special comment. Only two cases of the original thirty-four classifier predicates are identified as postpositions, which are tractable cases (see Section 3.4.2). With the rest of the classifier predicates, the ordinary translation of their corresponding Spanish expressions into sequences of signs should be valid for the purpose of communication. Almost all other classifier predicates found in the corpus could be generated by identifying some linguistic structure in the source language, e.g., '*el perro corre detrás de la pelota* (the dog runs after the ball)'. In this case, the classifier construction uses the spherical classifier handshape, representing the ball, and the quadruped classifier handshape, representing the dog. Both hands are moved: the 'dog's hand' after the 'ball's hand'. Another, much easier example is '*el perro está bajo la mesa* (the dog is under the table)', translated by placing the dog's hand under the table's hand. As can be noted, both examples present spatial movements or relations between objects. However, consider the example in (30) containing the expression '*no pueden dormir* (they cannot

sleep)' which has been translated by a classifier construction with the meaning of 'person tossing and turning into the bed'.

(30) a. *Muchas veces por la noche no pueden dormir preocupados por ese problema*
'Many times they cannot sleep because they are worried about that problem'

b. MUCHAS-VECES POR-LA-NOCHE CLS: 2 "persona dando vueltas en la cama" PRE-OCUPAR+ PROBLEMA PRO.3
many-times at-night a-person-tossing-and-turning-into-the-bed to-worry-ASP.-CONT problem that

This construction could be performed iconically, using the passive hand to represent the bed, with the bed's classifier handshape, while the active hand represents the person, using the 'V' handshape and performing an alternative rotation over the passive one. It is important to note that world knowledge is required for the generation of this expression: sleeping takes place in a bed; when a person cannot sleep, s/he usually tosses and turns in the bed; this action is performed on a particular axis of rotation, etc. However, an alternative valid translation using signs could be accomplished with 'ELLOS DORMIR NO (lit. they to-sleep no)'. Other interesting translation examples can be found, as for '*canibalismo* (cannibalism)', in which the signer has created a two-handed classifier predicate simulating the grasping of different parts of his/her own body and the eating of them. The construction is marked with 'CLC'. However, the construction is similar to the sequence of signs: 'BRAZO COMER / CUERPO COMER / PIERNA COMER (arm to-eat, body to-eat, leg to-eat)'. In order to generate classifier constructions, a machine translation system should first identify them and then be able to generate a correct description. Due to the difficulty involved in the generation of such constructions, very few works on this subject are found in the literature — only Huenerfauth (2006) for ASL and López-Colino and Colás (2011) for LSE.

Chapter 4

Spanish Sign Language Computational Inflectional Morphology

According to Beesley and Karttunen (2003), all known phonological and morphological processes can be treated as regular relations definable in terms of regular expressions and they can be compiled into finite-state transducers (FSTs). The nature of the phonological form in sign languages is different from words, signs are made of a different material, they are not sounds, they are handshapes, orientations, locations, movements in space and non-manual components (NMCs). An appropriate way to represent movements should be something different to symbols, like continuous functions. However, HamNoSys (Hamburg Sign Language Notation System) representations will be used to represent the phonology of a sign. The proposed model could be elegantly extended to cover other phenomena within Spanish Sign Language (LSE) as well as to other sign languages. The transducer resulting from the compilation of rewrite rules can be applied to morphological sign generation in machine translation or automatic sign recognizing through the analysis of HamNoSys transcriptions.

4.1 Phonology

Sign language words are signs, defined as articulatory structures performed mainly with the hands. The sign structure is described in most sign languages with four articulatory parameters or cheremes equivalent to the phonemes of oral languages so that their values have been found meaningful in a phonological contrastive analysis. These parameters are whether the sign in question is made with one hand or two hands and which is the symmetry in the latter case (S), handshape (H), orientation (O), location (L) and

movement (M). However, for LSE, it is also considered an articulatory parameter whether or not there is contact between the hand and the body in signs performed at a body location.[1] With the previous articulatory model in mind, the performance of one-handed signs is described as follows: the active or dominant hand is positioned at the location, contacting or not with that location if it is a body location, then the hand takes a handshape and an orientation, and finally, the hand performs a movement. In the case of two-handed signs, it can be distinguished asymmetrical two-handed signs with a passive hand, which acts as the location of the sign, and two-handed active signs, in which hands have a similar movement that can be symmetrical or anti-symmetrical.

Other non-manual components are superimposed to the manual articulation and, in most cases, they have a suprasegmental value. Non-manual components for LSE are given by the facial and the corporal expression. It has been identified as relevant the lips movements, caused by partial vocalizations and other mouthing patterns, the eyes movement, the elbow positions and the head and shoulder positions. Despite some scholars consider these components as phonological parameters, having the same role as stress or tone in spoken languages, in LSE, they play a principal role in morphology and syntax (Herrero Blanco, 2009).

4.2 HamNoSys Transcriptions of Signs

Writing systems differ from transcription systems in their purpose. While writing systems are intended for the purpose of communication, transcription systems are used for specialists to record or analyse the form of utterances. Transcription does not represent all the mechanical details of speech or signing but only those ones thought to be significant. A faithful reproduction of an utterance is only obtained by audio recording, in the case of spoken utterances, or by video recording, in the case of signed productions.

English writing and SignWriting are examples of writing systems, but IPA (International Phonetic Alphabet) and HamNoSys are transcription systems. With respect to IPA, it is worth noting that it is valid for transcribing languages without any written form. SignWriting was developed by Sutton (1981) inspired in DanceWriting, another system of the author created for writing down dance. SignWriting is a non-linear two-dimensional system where symbols are iconic and have internal structure: the signer's head is represented in SignWriting with a circle symbol and different eye shapes, eyebrows and mouth symbols are iconically arranged inside the head's symbol. HamNoSys is a linear transcription system for sign languages developed at the University of Hamburg (Prillwitz et

[1] A psycholinguistic study on the role of the above-mentioned phonological parameters in the processing of signs in LSE can be found in Gutiérrez-Sigut and Carreiras-Valiña (2009).

al., 1989). HamNoSys, as IPA, was designed as a phonological or phonetic transcription system. In the case of HamNoSys, symbols are highly iconic, since transcription systems emphasise writing speed and bundle different features. LSE has its transcription system called SEA (*Sistema de escritura alfabética*). The system uses the Latin alphabet and was designed for being easily employed with computers as a writing system (Herrero Blanco, 2003).

From the first version defined in 1984 and published by Prillwitz et al. (1989), HamNoSys has evolved to its current version 4 (Schmaling and Hanke, 2001). This latest version addresses issues arisen from to the use of HamNoSys in sign language generation and adds a new encoding scheme for the non-manual component. The HamNoSys notation for signs consists of a symmetry operator, an initial configuration and actions. In turn, the initial configuration consists of handshape, hand orientation and location. The HamNoSys encoding scheme, where mandatory parts are highlighting in bold, is the following:

$$\text{Symmetry} + \overbrace{\textbf{Handshape} + \textbf{Orientation} + \text{Location}}^{\textbf{Configuration}} + \text{Actions}$$

A handshape is specified by combining one of the basic base forms (, , , , , , , , , , ,) with diacritics for thumb position and finger bending, resulting in handshapes like . In addition, individual fingers involved as well as intermediate handshape forms can be described as well.

Hand orientation specification is composed of extended finger direction and palm orientation. Extended finger direction determines the tree-dimensional orientation of the hand using three different perspectives: the signer's view, bird's view and the view from the right. The palm orientation is relative to the extended finger direction. Moreover, adding a subscript to palm orientation makes palms orientation relative to the movement, changing as the direction of movement changes.

Location is specified by giving its position in the frontal plane and an optional distance from the body. If no location is specified, it is assumed the signing to take place in the 'neutral space', i.e., in a 'natural' distance from the upper torso. In two-handed signs where hands are related to each other, instead of using references to body, the 'hand constellation' is specified relatively, using one hand as a position for the other hand.

Actions (or Movements) combine path movements changing the hands positions, in-place movements of the hands and non-manual movements. The combination of these movements can be sequential o cotemporal. Basic types for path movements are straight lines, circles, curved and zigzag lines. In-place movements comprise changes to the handshapes or orientations of the hand, as well as other movements such as wiggling or twisting.

Additionally, diacritic symbols specifying the length of the movement can be added to basic movements and also a mode or manner of movement. HamNoSys indicates movement modality after the movement itself. The values considered are: fast (*), slow (⁻), tense (✖), hold at start (⁷) and sudden stop at end (‖). Note that the final location of the movement is determined by the direction and size of the movement path. However, in HamNoSys, the movement can be also specified by giving the target location.

Repetition can be added to movements. There are several symbols or combinations of symbols for basic repetition: repeated once (+), repeated twice (++), repeated a couple of times (⧺), repeated once with the second iteration starting where first ended (⧻), repeated a couple of times with each iteration starting where previous ended (⧼⧻⧽) and repeated in reverse direction (⟲). The symbol ~ can appear at the end of the movement transcription indicating a repetition in which hands swap roles. This marker can be added twice (~~) to indicate another repetition.

Two-handed signs can be symmetrical or asymmetrical. In symmetrical signs only the dominant hand, which is usually the right hand, is transcribed and it is the symmetry operator that determines how the description is copied to the non-dominant hand. There exists an inventory of eight symbols for specifying symmetry: mirror (¨), parallel (ˈ), etc. Exceptions are described by means of the ↙ symbol using the syntax '[*dominant* ↙ *non-dominant*]'. Asymmetrical two-handed signs are transcribed with no symmetry operator, parameter by parameter using the previous syntax when values for hands differ. In two-handed signs with a passive hand, the absence of movement in the non-dominant hand is noted by ⦰ in '[…↙⦰]'. This expression can also be used in symmetrical signs when, as an exception to symmetry, the non-dominant hand has no movement.

The different approaches to the encoding of non-manual aspects of sign languages are discussed in Hanke, Langer and Metzger (2001). Most of the recent approaches, including HamNoSys, prefer multi-tier representations to linear encodings for signed utterances. In the case of HamNoSys, also coding conventions for synchronisation of non-manuals information at sign-level are provided. However, in the actions part, each movement has the hands as default articulators. Other tiers that are synchronized to the master tier containing the manual part of the sign are the shoulders, body, head, gaze, facial expression and mouthing. Facial expression comprises of eyebrows, eyelids and nose. Except for mouthing, the rest of NMCs use an inventory of both static and dynamic descriptions as 'UL' for left shoulder raised or 'NO' for nodding the head up and down. Mouthing can consist of mouth gestures as well as mouth pictures, i.e., the articulation of tongue and lips derived from spoken language. Mouth gestures are represented from an inventory of gestures using codes. Mouth pictures are described as visemes encoded using SAMPA (Speech Assessment Methods Phonetic Alphabet). Mouth gestures and pictures

are intermingled in one string.

HamNoSys is ambiguous in the sense that a sign can be expressed in a variety of forms from the point of view of performance. This ambiguity poses difficulties to automatic processing since this formal variation should be taken into account. However, without loss of generalization, in this thesis, signs will be transcribed following some conventions that ease morphological processing. These conventions will be presented when appropriate in further sections.

4.3 Morphology

Minimal units of meaning and grammatical values can be expressed as independent single words called free morphemes, or as additions or alterations to the phonology of words, known as bound morphemes. Sign languages have a wide range of morphological processes, but this chapter deals mainly with inflection and modulation in LSE, leaving aside derivation, compounding and classification. LSE, and virtually all sign languages studied, have a rich inflection system. Bound morphemes found in LSE belong to these two classes: flective and introflective. Flective morphemes are concatenated to the phonological form, like a movement or a repetition of the original movement, sometimes in a direction.[2] Introflective morphemes cause internal changes to the phonological form of the sign. Consequently, these morphemes are non-concatenative in nature.[3] LSE has in some cases irregularities as suppletive base forms, i.e., different and unrelated base forms for the same sign, indicating different grammatical values. Finally, modulation is the modification of suprasegmental elements, like the stress or the tone in spoken languages.

LSE has systematic morphemes, which are always realised in the same way, but also has variable morphemes. Degree morphemes in LSE belong to the second type of morphemes and they can be found in adjectives, verbs and adverbs. Their values are referred to as intensification and restriction. They are realised in a variety of forms with NMCs as the muscular tension, the length and speed of movement, etc.

In LSE, gender and other grammatical values are expressed by means of free morphemes, e.g., in the case of gender, by performing the signs 'HOMBRE (man)' or 'MUJER (woman)' after the sign. However, flective morphemes such as repetition are used to express several forms of plural in nouns, some aspectual values in verbs and predicative adjectives and in noun-verb derivation similar to the way described for American Sign Language (ASL) in Supalla and Newport (1978). In verbs, introflection changing hand-

[2]Note that, in the transcription system, repetition is a symbol added to the movement, but repetition can be seen such as reduplication, a non-concatenative operation.

[3]Introflective morphemes are also found in Semitic languages, which are known for having a root-and-pattern morphology.

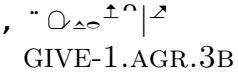
shape, orientation, location, or movement is used to agree with the subject, the object, the receiver or the locative. In nouns and pronouns, signing location is changed to mark the locative case. Also numeral incorporation takes place in verbs and pronouns by introflecting the numeral's handshape. Modulation is used to express aspect by means of NMCs as speed, tension and length of movement. A classic description of aspectual modulations in ASL can be found in Klima and Bellugi (1979). Adverbs are also modulated to express intensification and restriction. However, the possibility to express morphologically the aforementioned grammatical values is conditioned by phonological and semantic constraints, as it will be seen in the formalization of each phenomenon.

4.4 Previous Formalization of Sign Language Morphology

The phonological models of how signs are composed begin historically with Stokoe (1960) and continue with the Move-Hold model (Liddell and Johnson, 1989), which borrows some ideas from auto-segmental phonology, and with the Hand Tier model (Sandler, 1989), which attempts to improve the previous models. However, Johnston (1991b) advocates the use of interlinear transcriptions and glossing for presenting and discussing sign language data. All these sign language data are transcribed using HamNoSys. An inflected sign is represented by its citation form (stem form) plus its morphological variation. Variations are separated with bars and placed in order after the stem, as in the following example of the format for representing a gloss with two morphs added -MORPH1 and -MORPH2 changing, respectively, the handshape and the movement of the citation form:

(1) Handshape Orientation Location Movement | Handshape1 | Movement2
GLOSS-MORPH1-MORPH2

Some labels that capture the general function or meaning of morphemes are also given as temptative: -LOC, -ORI, -DIR, -AGR, -ASP, -INTENS, etc. In Johnston (1991a) the transcription system was put to work for the analysis of verbal signs in Auslan (Australian Sign Language), giving examples like the following:

(2) 
PRO.1 GIVE-1.AGR.3B
'I gave him (the beer)'

Much of the previous work done in sign language processing, including computational morphology, has been carried on within the ViSiCAST Project (Schulmeister, 2001). In

that project, Hanke (2002) showed how the morphological information given in the transcription of Johnston (1991b) could be abstracted using lambda calculus. Citacion forms in the lexicon are represented as lambda expressions like $\lambda x.$ ¨ ◯⍙ᵡ x ↘↓, where location has been abstracted. In general, a lambda expression can take location, source, classifier, etc. and, after reduction, a complete HamNoSys expression is obtained. However, due to the number of obligatory and optional morphemes that a sign can admit, many helper functions are needed, making lambda expressions difficult to interpret. The previous model based on lambda calculus was replaced by another model in which phonological information is split up into parameters and represented with feature structures like the following, corresponding to the German Sign Language sign HOUSE:

HANDEDNESS: ¨
HANDSHAPE: ◯
ORIENTATION: ⍙
HANDCONSTELLATION: ᵡ
LOCATION: ⊡
MOVEMENT: ↘↓

The previous is a simplified version of the kind of feature structures used, since they adopted the framework of the Head-driven Phrase Structure Grammar (HPSG; Pollard and Sag (1994)) in which linguistic information at different levels, including the phonological level, is coindexed. Also types are assigned to signs depending on their inflectional model. Figure 4.1 contains a part of the subsumption hierarchy modelling some sign types. For readability reasons, the subtype hierarchy corresponding to locatable signs, which implements different constraints on the locative possibilities of signs, has been collapsed under the type *locatable-sign*. In this framework, typed unification is used as the operation for composing morphological and lexical information, which is finally mapped into a HamNoSys sequence. The system deals with the morphological phenomena of numeral and classifier incorporation, directional verbs and locatable signs. Also in Elliott et al. (2001) plurals and the aspectual inflection are implemented.

Continuing with the work started in ViSiCAST, Marshall and Sáfár (2004) and Elliott et al. (2008) presented an HPSG sign language generation component for an English-to-BSL prototype system. The system deals with the allocation of positions in the signing space denoting nominals. Subsequent pointing at these locations (indexing) is the equivalent of pronominal reference in spoken languages. Locatable nominals can be signed at specific positions. Non-locatable nominals can be positioned by indexing a particular location after sign. Other nominals can be anaphorically referred to by the inclusion of their classifier handshapes within verbs of manipulation. Verbs are sub-classified into

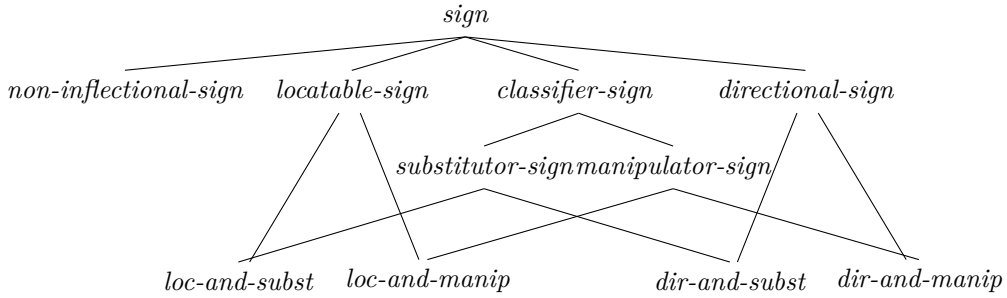


Figure 4.1: ViSiCAST simplified type hierarchy

fixed, directional (parametrised with start and end positions) and manipulative (parametrised by a proform realised with a classifier handshape). However, combinations of these verbal types are allowed, as for TAKE, which is a directional-manipulative verb. The BSL glossing of the sentence 'I take the mug' can be seen in (3). The synthesis of TAKE is achieved from 'px' and 'p1', which are the locations where 'mug' and 'I/me' will be signed. The orientation for TAKE is constrained to be consistent with the start and end positions and the handshape needed by TAKE is obtained from the manipulative classifier handshape of MUG.

(3) MUG(px) , TAKE(px,p1,manip_handshape(MUG)) , ME(p1)
'I take de mug'

Directional verbs must agree with the locations of their subject and object and in number with their arguments. However, for plural nominals, a group of individuals or objects can be positioned approximately at the same position of a single individual or object, but adding a displaced repeated motion, as in the example (4). In the example, the parametrised lexical entry for BOWL, is instantiated to $\langle \#\rightarrow \rangle$ in the plural case, otherwise the empty list is used for singular. Sign languages, like spoken languages, make a semantic distinction between collective (the entire set) and distributive (each member of the set) uses of the plural. A directional verb like PUT is represented as the list [ClassifierShape, SourcePosition, R1, Motion, DestPosition, R2], which admits a modification for the agreement with a plural collective reading consisting of a sweeping motion prior to the movement. This is achieved by instantiating R1 with the appropriate HamNoSys subsequence, or by the repetition of the movement for a distributive reading, instantiated in R2. In example 4, SourcePosition and ClassifierShape are taken from BOWL and DestPosition and orientation are taken from SINK. The direction of motion is made consistent with the start and end positions. As it can be seen R1, is instantiated with the empty list and R2 with a repeated displaced movement.

(4) 
BOWL(px,pl) SINK(py) PUT(px,py,classifier(BOWL))
'I put the bowls in the sink'

Shield and Baldridge (2009) present a different approach consisting of a finite-state morphological analyser for modelling some of the verbal aspectual distinctions in ASL. In their approach, signs are not represented by transcriptions in any formalism, but by a more abstract formal representation not committed to a particular underlying phonological form of the sign. In this representation, each parameter is represented with a transparent and comprehensible shorthand system. They propose the following parameters and contrastive values, which capture the morpho-phonological information needed to address the problem of modelling verbal aspect in ASL:

- Types:
  - 1H: One-handed
  - 2HS: Two-Handed Symmetrical
  - 2HF: Two-Handed Dominant
- Handshapes: A, B, C, 5, E, F, G, H, 3, O, R, V, W, X, Y, 8
- Locations: face, neutral, torso, neck, shoulders, chest, trunk, upper arm, elbow, forearm, wrist
- Palm orientations: up, down, out, in, base
- Movements: ±touch, ±twist, ±reduplication, ±arc, ±slow

Using the previous notation, the ASL verb SEE is represented as follows:

<Type:1H DH:Vin NDH:none Loc:face −Touch −Twist −Redup +Arc −Slow>

indicating that it is one-handed (Type:1H) with a dominant hand (DH) with a handshape with the index and middle fingers extended (V) and the palm oriented towards the signer (in). The movement begins at the height of the face (Loc:face) without touching it and the movement continues with a normal speed (−Slow) describing an arc (+Arc). In ASL, as in other sign languages, aspectual distinctions mainly affect to movement. Only the two most common aspectual values, habitual and continuative, are considered in the paper. The following is a rewrite rule for the habitual aspect:

(5) (− → + / .*Aspect:Hab.* ___ Redup|Arc)

In Shield and Baldridge (2009), some unpronounceable verbal forms are generated when some aspect values are added. These co-articulation constraints are dealt with by a series of transducers composed together in a cascade correcting the forms appropriately.

4.5 A Formalisation Proposal of LSE Morphology

Word formation and morphological alternations are central aspects of morphology. The former, word formation, states the principles governing the combination of stems and affixes. The latter, morphological alternations, accounts for the different realisations of a morpheme depending on its phonological environment and the morphological composition of the word.

In a computational morphology, one wants to relate the lexical level, consisting of baseforms with abstract linguistic descriptions in the form of feature sets to surface realisations. Chomsky and Halle (1968) formalized traditional phonological grammars using ordered sequences of rewrite rules transforming abstract phonological representations into surface forms through a series of intermediate representations. Such rewrite rules have the general form ($\alpha \rightarrow \beta \ / \ \gamma \ \_\_\_ \ \delta$) where α, β, γ and δ are arbitrarily complex strings or feature-matrices. Johnson (1972) established that such rewrite rules describe regular relations when α, β, γ and δ are regular expressions. Regular relations can be implemented as finite-state transducers. As a consequence, an ordered series or cascade of rules can be compiled into FSTs and composed to obtain a single FST representing the entire rule series. These results were rediscovered and put into a more algebraic form by Kaplan and Kay (1994), who developed the techniques for the compilation of rewrite rules to FSTs. According to Beesley and Karttunen (2003), all known phonological and morphological processes, organized by different theories in different ways, can be treated as regular relations definable in terms of regular expressions. Regular relations are compiled into finite-state transducers, which are bi-directional and can be used both for analysis and generation.

Lexical baseforms or lemmas are conventional forms that usually coincide with dictionary headwords. However, these conventions are not universal. For some languages like Spanish, infinitive forms are chosen as verbal lemmas, while in others like Latin, the first-person singular present-indicative active forms are used instead. In languages such as Sanskrit or Esperanto stems are used as baseforms instead of full forms, but for Semitic languages, due to its non-concatenative morphology, roots formed by a triplet of consonants are typically used as headwords. For a discussion on the notion of lemma in several languages, see Knowles and Mohd Don (2004). In sign language dictionaries, entries can be accessed via glosses, since almost all of dictionaries are bilingual, or by specifying their phonological parameters. Many sign language dictionaries contain drawings, photographs and, with the advent of electronic dictionaries, video performances. When signs inflect, its lexical baseform correspond to a form with prototypical morphemes, which is considered neutral. Signs in the DILSE-III (Fundación CNSE, 2008) can be searched by parameter using a SignWriting-like description. In addition to the video performance, a

SEA transcription of the signs is also given.

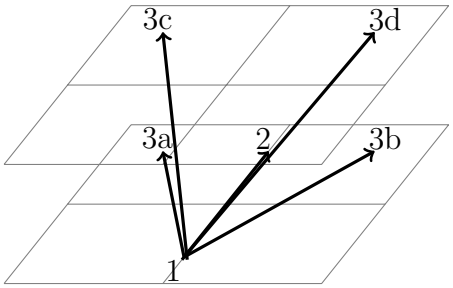
Lexicon entries are classified according to their inflectional paradigm, i.e., the way the baseform should be manipulated in order to generate all the possible morphologically-related sign forms. At the lexical level, a sign's baseform is phonologically represented with its HamNoSys transcription and its morphological features, which are represented by symbols like '-SG' indicating 'singular' number value. These features are given in a fixed order. At surface level, the sign's wordform is only represented by its HamNoSys transcription.

The notation used here for regular expressions is based on POSIX-like regular expressions. Most characters or symbols are treated as literals, i.e., they match themselves, but there are some metacharacters, i.e., characters with a special meaning. A dot '.' matches any single symbol. The caret '$\widehat{\ }$' and the dollar '$' match, respectively, the starting position and the ending position of the string. Parentheses are used for grouping expressions. The metacharacter '$^+$' is a postfix operator that matches the preceding regular expression one or more times. A '?' makes optional the preceding expression. There is another repetition operator, represented by '*' matching its operand zero or more times. A vertical bar '|' is disjunction operator matching any of the expressions on both sides. Precedence and associativity of operators are defined as in the standard. Any metacharacter escaped with a backslash '\' is interpreted literally. Finally, concatenation is expressed by juxtaposition of regular expressions and 'ϵ' denotes the empty string. For regular relations, the operator '$\circ$' is used for the composition of two rational expressions. Direct rewrite rules and conditional rewrite rules have the syntax $(\alpha \rightarrow \beta)$ and $(\alpha \rightarrow \beta \ / \ \gamma \ \_\_\_ \ \delta)$ respectively, being all the operands regular expressions. The application of rewrite rules is meant to be left-to-right with a longest match strategy.

Let us define the alphabet Σ as the union of the set of HamNoSys symbols and the set of morphological tags and let be S, H, L, O and M regular languages over Σ describing each of the string values of each of the phonological parameters of signs. Let be *Tags* the set of non-empty sequences of morphological tags. To give an example, in the case orientation O, its language is defined by the following set of regular expression definitions:

FingerBaseOrientation =
FingerOrientation = *FingerBaseOrientation FingerBaseOrientation*?
PalmOrientation =
O = *FingerOrientation PalmOrientation*

In Wintner (2007) the use of finite-state technology in large-scale morphological grammar development was analysed and two main problems were identified: abstraction and



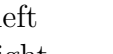
Points	Side	Proximity	$O + L$
1	central	near	
2	central	near	
3a	left	near	
3b	right	near	
3c	left	far	
3d	right	far	

Figure 4.2: Grammaticalised points in space and their expression as combination of orientation (O) and location (L) HamNoSys parameter values

incremental development. In order to address the problem of abstraction in the LSE morphology, rules are designed inspired in operational semantics of programming languages as series of steps, implemented with simpler rules that ultimately perform simple rewritings of the phonological representation of a base sign. This modular design reuses many rules making developing of new rules, testing, debugging and maintenance less difficult and costly.

4.6 Deictics

Deixis is a grammatical phenomenon by which persons, objects, space and time relative to the context of an utterance are referenced by means of personal pronouns, demonstratives, locatives and temporals. In sign languages, pronominal systems may seem at first glance to be uniform due to the analogue use of space (McBurney, 2002), but they display as much variation as in spoken languages. In LSE, deictics are based on three positions expressing proximity. The third position value splits into proximal/present and distant/absent, and those values are split again into left/right locations. These points and how their values are realised in orientation and location parameters can be seen in Figure 4.2.

In LSE, temporal deictics are expressed with invariable signs. However, personals, demonstratives and locatives are morphologically related, since all singular values are realised with the upright index. Personal pronouns do not include movement for singular, but demonstratives include a pointing movement and locatives a directional movement. Personals and demonstratives admit unspecific plural quantification by means of inflection and numeral incorporation up to five by introflection. Generic plurals add to their singular base form an internal wrist movement while specific plurals incorporate numerals up to five, by substituting the handshape of the base sign by the numeral's handshape. Additionally, in specific plurals, an oscillating movement is added and the orientation changes from pronation to supination.

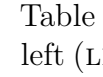
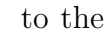
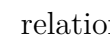
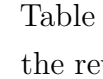
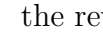
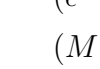
Tags	-SG	-PL	-PL.2	-PL.3	-PL.4	-PL.5
-1ST						
-2ND						
-3RD-LEFT-PROX						
-3RD-LEFT-DIST						
-3RD-RIGHT-PROX						
-3RD-RIGHT-DIST						

Table 4.1: LSE personal pronouns. Third person (3RD) pronouns are split according to left (LEFT) or right (RIGHT) location and proximal (PROX) or distal (DIST) distance values. Columns PL.2 to PL.5 contains pronouns with incorporated numerals.

Example (6) shows the personal pronoun and the demonstrative signs corresponding to the third person, dual plural, left located and distant forms. It is shown the regular relations associating phonetical forms their lexical features that include category, subcategory, person, number, lateralization and distance from the signer.

(6) a.  (PRON-PERS-3RD-PL.2-LEFT-DIST)

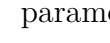
b.  (PRON-DEM-3RD-PL.2-LEFT-DIST)

Table 4.1 contains the complete paradigm for personal pronouns and rule (*4.1*) some of the rewritings used, in particular, the rules dealing with the formation of dual plurals:

DeicticsInflection =

$$
\begin{aligned}
&\ldots \\
&(H \rightarrow \text{PRON} \;/\; \widehat{\ } \;\_\_\_\; O \;.^*\; (\text{-PERS}|\text{-DEM})\; .^*\; \text{-PL.2}) \circ \\
&(F \rightarrow \text{INDEX} \;/\; \widehat{\ }H \;\_\_\_\; P \;.^*\; (\text{-PERS}|\text{-DEM})\; \text{-3RD}\; \textit{PL.N}\; .^*\; \text{-DIST}) \circ \\
&(P \rightarrow \text{PALM} \;/\; \widehat{\ }\; H\; F \;\_\_\_\; L\; .^*(\text{-PERS}|\text{-DEM})\; \textit{NON.1ST}\; \textit{PL.N}) \circ \\
&(L \rightarrow \text{CHEST} \;/\; O \;\_\_\_\; M\; (\text{-PERS}|\text{-DEM}|\text{-LOC})\; .^*\; \text{-DIST}) \circ \\
&(\epsilon \rightarrow \text{RIGHT} \;/\; O \;\_\_\_\; L\; M\; (\text{-PERS}|\text{-DEM}|\text{-LOC})\; .^*\; \text{-LEFT}) \circ \\
&(M \rightarrow \text{PL.2} \;/\; L \;\_\_\_\; \text{-PERS}.^*\; \textit{NON.SG}) \circ \\
&\ldots
\end{aligned}
\tag{4.1}
$$

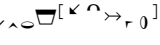
Note that there is no specific base form here associated to the sign PRON, which initially enables all the phonological parameters (H, F, P, L and M), where orientation O has been decomposed into F and P, the finger and palm orientations. The values of these parameters are subsequently constrained by a set of rewriting rules applied in cascade using composition. Some regular expressions can be defined for convenience or in order to organize the morphological information, as non-first person values *NON.1ST* = -2ND|-3RD

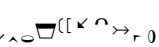
or specific plurals *PL.N* = -PL.2|-PL.3|-PL.4|-PL.5. In the following, PERS.1, PERS.3A, etc. will be used in examples as shorthand for the singular personal pronoun signs PRON-PERS-1ST-SG, PRON-PERS-3RD-SG, etc.

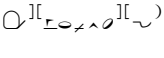
4.7 Nominal Inflective Plurals

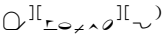
In LSE, numerals or indefinites are added after singular nouns to express number, as in the examples in 'HERMANO TRES (three brothers)' and 'COCHE ALGUNO (some cars)'. Although the sign 'MUCHOS (many)' can be added after a sign to express plural, repetition is the most iconic representation of number and, in almost all sign languages, nominal signs and classifiers can add some kind of repetition to their singular form to express frequency, abundance or descriptive plurals. The use of repetition in nouns for number inflection depends on semantic and phonological constraints (Fernández-Soneira, 2008; Herrero Blanco and Peidro, 2007). In order to admit repetition, a sign should be one-handed and signed in the neutral space, two-handed symmetrical or two-handed with a passive hand. Body-anchored signs and many signs including repetition do not use to inflect for number. Normally, adding a repetition to a sign means to perform the sign twice, but when repetition is included in the sign's movement, the sign should be performed three times. As it will be seen in this section, number inflection has many variations. In order to implement them, the following morphematic tags, which account for the different realisations of number, are proposed: PL.REP (in-place repetition), PL.REP.C (in-place repetition with circular joining movement), PL.REP.H (horizontal directional repetition), PL.REP.V (vertical directional repetition), PL.SLID.H (horizontal directional sliding), PL.SLID.V (vertical directional sliding), PL.BIMAN.H (two-handed repetition with alternating horizontal movement) and PL.BIMAN.V (two-handed repetition with alternating vertical movement).

Signs like 'NOCHE (night)' or 'SÍMBOLO (symbol)' express abundance using in-place repetition, as it can be seen respectively in examples (7) and (8), but repetition is applied differently. For NOCHE, which is a two-handed symmetrical sign, repetition affects the performance of the two hands, but for SÍMBOLO, which is a two-handed asymmetrical sign with a passive hand, repetition applies only to the active hand.

(7) a. 
(NOCHE-SG)

b. 
(NOCHE-PL.REP)

(8) a. 
(SÍMBOLO-SG)

b. 
(SÍMBOLO-PL.REP)

The rule *RepeatInPlace* (*4.2*) implements the in-place repetition of signs dealing with one-handed and two-handed signs. It first applies the rule *MarkMovement* (*4.3*) that

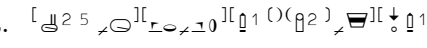
groups movement symbols and marks the point of insertion of the repetition operator (>). This marking depends on how hands are used. Only the movements of the active hands are marked. Once the movement is marked, *InsertInPlaceRepetition* (*4.4*) replaces the marking symbol by a repetition symbol.

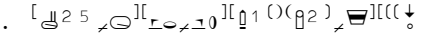
$$RepeatInPlace = MarkMovement \circ InsertInPlaceRepetition \tag{4.2}$$

$$\begin{aligned} &MarkMovement = \\ &\quad (\epsilon \rightarrow {}^{\lceil} \;/\; L \_\_\_ M) \circ (\epsilon \rightarrow {}^{\rfloor}> \;/\; L^{\lceil}M \_\_\_ Tags) \circ \\ &\quad (\epsilon \rightarrow {}^{\lceil} \;/\; {}^{[} \_\_\_ M_{\curvearrowright}) \circ (\epsilon \rightarrow {}^{\rfloor}> \;/\; L^{[\lceil}M \_\_\_ {}_{\curvearrowright}) \end{aligned} \tag{4.3}$$

$$InsertInPlaceRepetition = (> \rightarrow {}^{+}) \tag{4.4}$$

Other signs express their frequency plural by joining each of the in-place repetitions by means of a circular movement, like 'LUNES (Monday)', as can be seen in the example (9b).

(9) a.  (LUNES-SG)

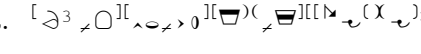
b.  (LUNES-PL.REP.C)

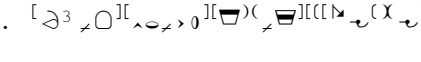
Rules dealing with this kind of repetition are:

$$RepeatCircular = MarkMovement \circ InsertInPlaceCircularRepetition \tag{4.5}$$

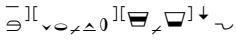
$$InsertInPlaceCircularRepetition = (> \rightarrow {}^{\circlearrowleft}) \circ MarkMovement \circ (> \rightarrow {}^{+}) \tag{4.6}$$

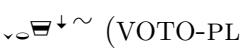
Many nouns that are repeated to express frequency are deverbal nouns. Consequently, their plural forms do not differ from the corresponding verbal signs displaying a frequentative aspect value. Two examples of deverbal nouns in LSE are 'CARTA (letter)' and 'VOTO (vote)'. The first case, CARTA is derived from the verbal sign with the meaning of 'sending letters' and is repeated in-place, as in example (10).

(10) a.  (CARTA-SG)

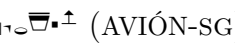
b.  (CARTA-PL.REP.C)

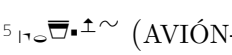
However, as can be noted in example (11), VOTO-SG is articulated as a two-handed asymmetrical sign with the passive hand representing iconically an urn, but its frequentative plural which is realised using both hands as articulators, alternating their movement when the sign is repeated.

(11) a.  (VOTO-SG)

b.  (VOTO-PL.BIMAN.V)

Also one-handed signs, as 'AVIÓN (airplane)', turns two-handed to express its plural, but unlike VOTO, which is symmetrized in the vertical plane, the symmetry for AVIÓN is horizontal, see example (12b).

(12) a.  (AVIÓN-SG)

b.  (AVIÓN-PL.BIMAN.H)

The following rewrite rules are applied to singular base forms and making signs two-handed symmetrical alternating hands horizontally in repetition. Because the final location of the movement can be determined in HamNoSys either by the direction and size of the movement path or by specifying it as the target location, the parameter M is decomposed as $M = RT$, where R represents all the relative movements and T the target locations. Note that absolute movements specifying hands as target locations are removed with rule (*4.8*).

$$\begin{aligned} &\mathit{MakeAlternatingHandsHorizontally} = \\ &\quad \mathit{RemoveTargetLocation} \circ \\ &\quad \mathit{ProjectDominantHand} \circ \\ &\quad \mathit{InsertParallelSymmetry} \circ \\ &\quad \mathit{MarkMovement} \circ \\ &\quad \mathit{InsertAlternatingMovement} \end{aligned} \tag{4.7}$$

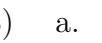
$$\mathit{RemoveTargetLocation} = (T \rightarrow \epsilon \;/\; {}^{\lceil}R \;\_\_\_\; {}_{\lrcorner}M^{\rceil}) \tag{4.8}$$

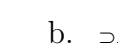
$$\mathit{InsertParallelSymmetry} = (\epsilon \rightarrow \; ^{:} \;/\; ^{\frown} \;\_\_\_\;) \tag{4.9}$$

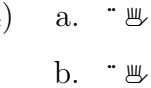
$$\mathit{InsertAlternatingMovement} = (> \rightarrow {}^{\sim}) \tag{4.10}$$

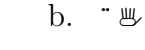
$$
\begin{aligned}
&\textit{ProjectDominantHand} = \\
&\quad (S \rightarrow \epsilon) \circ \\
&\quad (H|O|L|M \rightarrow \epsilon \;/\; \curlywedge \_\_\_\_\; ^{]}) \circ \\
&\quad (^{[} \rightarrow \epsilon \;/\; \_\_\_\_\; (H|O|L|M)_{\curlywedge}^{]}) \circ \\
&\quad (_{\curlywedge}^{]} \rightarrow \epsilon)
\end{aligned}
\tag{4.11}
$$

Directional repetition is used to express abundance plurals and is applied to objects represented mentally as spatial-ordered series. Default direction of repetition is horizontal, going from left to right, as it can be seen for 'PERSONA (person)' in example (13), but in other cases, direction is vertical, from the height of the baseform to the bottom, as it can be seen in example (14) for 'ORACIÓN (sentence)'.

(13) a.  (PERSONA-SG)

b.  (PERSONA-PL.REP.DIR.H)

(14) a.  (ORACIÓN-SG)

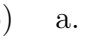
b.  (ORACIÓN-PL.REP.DIR.V)

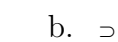
Horizontal or vertical directional repetition is orthogonal to the sign's own direction. Horizontal repetition is realised with the following rules and vertical repetition is realised in a similar way.

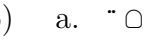
$$
\textit{RepeatHorizontally} = \textit{MarkMovement} \circ \textit{InsertHorizontalRepetition} \tag{4.12}
$$

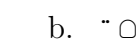
$$
\textit{InsertHorizontalRepetition} = (> \rightarrow {}^{(+\rightarrow)}) \tag{4.13}
$$

Sometimes, directional repetition becomes a sliding movement, like in the plurals of 'ÁRBOL (tree)', 'CASA (house)' or 'PALABRA (word)', which can be seen in the following examples:

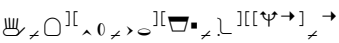
(15) a.  (PALABRA-SG)

b.  (PALABRA-PL.SLID.H)

(16) a.  (CASA-SG)

b.  (CASA-PL.SLID.H)

(17) a. 
(ÁRBOL-SG)

b. 
(ÁRBOL-PL.SLID.H)

Horizontal sliding is realised with the following rules:

$$SlideHorizontally = InsertHorizontalSlide \tag{4.14}$$

$$\begin{aligned} &InsertHorizontalSlide = \\ &\quad (\epsilon \rightarrow {}^{\rightarrow} \;/\; {}^{[}M \;\_\_\_\; {}_{\curvearrowright}) \circ (\epsilon \rightarrow {}^{\rightarrow} \;/\; {}_{\curvearrowright}M \;\_\_\_\; {}^{]}) \circ (\epsilon \rightarrow {}^{\rightarrow} \;/\; M \;\_\_\_\; Tags) \end{aligned} \tag{4.15}$$

Additionally, the plural of abundance can adopt a form of a distributive plural. In this kind of plural, the sign is repeated several times and each instance is placed in a particular location, as in 'ÁRBOL-LOC.LEFT ÁRBOL-LOC.CENT ÁRBOL-LOC.RIGHT'. The lateralisation of the sign, which should be locatable, is changed with rewrite rules as the following, which takes into account when the sign is two-handed and the passive hand serves as a location to the active hand:

$$\begin{aligned} &MakeLocationRight = \\ &\quad RemoveLateralisationSymbolFromLocation \circ \\ &\quad (\epsilon \rightarrow \blacksquare \;/\; L \;\_\_\_\; {}^{[}?M?\; Tags^{*}\text{-LOC.RIGHT}) \circ \\ &\quad (\epsilon \rightarrow \blacksquare \;/\; {}^{[}L_{\curvearrowright}L \;\_\_\_\; {}^{]}.^{*}\text{-LOC.RIGHT}) \end{aligned} \tag{4.16}$$

Finally, when repetition is used to emulate the arrangement of objects in the real world, using the signs or their classifiers, it is said that the plural is a descriptive plural. This kind of plural is not considered a proper morphological form of plural. It is worth noting that there exist a continuum between systematic or morphological plurals and no systematic or descriptive plurals, as is evidenced by directional plurals, which are systematic and descriptive at the same time.

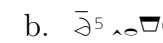
4.8 Nominal Introflective and Suppletive Plurals

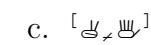
A few temporal nouns in LSE admit numeral incorporation introflecting the configuration of the number into the sign. These nouns correspond to the names of time units: 'MINUTO (minute)', 'HORA (hour)', 'DÍA (day)', 'SEMANA (week)', 'MES (month)' and 'AÑO (year)'. They all admit numerals up to ten that are represented in the configuration. In addition,

when numerals are incorporated, some other phonological parameters of the sign are modified, so that, in some cases, introflection takes place in a suppletive form of the sign. Suppletive forms of plural exist in some LSE communities or in some registers for the temporal nouns HORA, MES and SEMANA. Suppletion supposes the substitution of the neutral base form of the sign by another base form conveying some grammatical values, which, in the case of temporal nouns, is the plural value.

In order to show how suppletion and inflection interact, let us suppose an LSE variety in which DÍA and AÑO have suppletive forms of plural. In the case of DÍA, the singular form, glossed by the sign DÍA-1, and the incorporation of three, glossed by DÍA-3, are shown in example (18). For AÑO, examples are given in (19).

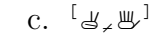
(18) a.  (DÍA-1)

b.  (DÍA-3)

c.  (DÍA-7)

(19) a.  (AÑO-1)

b.  (AÑO-2)

c.  (AÑO-7)

The replacement of the phonological base form in the case of signs with suppletive plural is realised by the rules in (*4.17*), which implements a list of signs with its corresponding plural suppletive form.

SuppletivePlurals =
($AÑO \rightarrow AÑOS$ / ⌢ ___ (-2|-3| . . . |-10)) ∘
($DÍA \rightarrow DÍAS$ / ⌢ ___ (-2|-3| . . . |-10)) (*4.17*)

For the incorporation of numerals above five, its realisation is assumed to be parallel to the construction of numerals. The configuration of 'CINCO (five)' is incorporated into the passive hand and the active hand incorporates the numeral resulting of the subtraction of five from the original numeral. Rules in (*4.18*), (*4.19*) and (*4.20*) are responsible for the incorporation of numerals into one-handed signs.

IncorporateNumeral =
IncorporateNumeralToDominantHand ∘
IncorporateNumeralToNonDominantHand (*4.18*)

IncorporateNumeralToDominantHand =

$$(H \rightarrow \text{☝} \;/\; \widehat{\ }S? \;\_\_\_\; O.^*(\text{-}2|\text{-}7)) \circ (H \rightarrow \text{🖐}^5 \;/\; \widehat{\ }S? \;\_\_\_\; O.^*(\text{-}3|\text{-}8)) \circ$$
$$(H \rightarrow \text{✌} \;/\; \widehat{\ }S? \;\_\_\_\; O.^*(\text{-}4|\text{-}9)) \circ (H \rightarrow \text{🤟} \;/\; \widehat{\ }S? \;\_\_\_\; O.^*(\text{-}5|\text{-}10)) \qquad (4.19)$$

IncorporateNumeralToNonDominantHand =

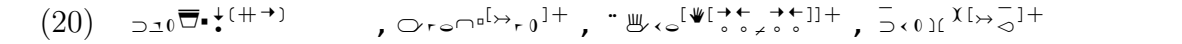
$$(\epsilon \rightarrow \text{🤙} \;/\; \widehat{\ }S? \;\_\_\_\; H.^*(\text{-}6|\ldots|\text{-}10)) \circ (\epsilon \rightarrow \text{🤙}^{\text{🤟}} \;/\; \text{🤙}H \;\_\_\_\; O.^*(\text{-}6|\ldots|\text{-}10)) \circ$$
$$(\epsilon \rightarrow \text{🤙} \;/\; O \;\_\_\_\; L\ .^*\ (\text{-}6|\ldots|\text{-}10)) \circ (\epsilon \rightarrow \text{🤙}^{\text{▪}} \;/\; \text{🤙}L \;\_\_\_\; M?\ (\text{-}6|\ldots|\text{-}10)) \qquad (4.20)$$

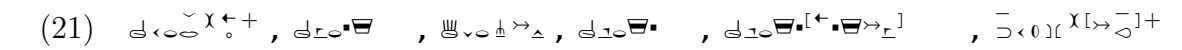
4.9 Verbal Agreement

From a morphological point of view, verbs are usually classified according to the way they agree with their arguments. In general, it can be said that there exists some degree of iconicity in the performance of the agreement, since subject agreement precedes object agreement and object agreement precedes recipient agreement. Let us consider the example of 'DAR (to give)', which is a ditransitive verb. Subject agreement is expressed in L, object in H and recipient and action are expressed in M. However, one can find signs showing any combination of assignments between grammatical functions and phonological parameters. In addition, manual agreement can be accompanied by NMCs, or even be substituted, especially using the trunk and head positions or the gaze.

According to Padden (1990), verbs in sign languages can be morphosyntactically classified into plain verbs, agreeing verbs and spatial verbs. However, the borders between these classes are not crisp. Another alternative classification of verbs reflecting morphological and semantic differences was proposed by Engberg-Pedersen (1993). In her classification, there is a continuum between what she called non-polymorphemic and polymorphemic verbs having, respectively, plain verbs and spatial verbs as both ends. In LSE verbs are classified into four groups: plain, classificatory, directional and spatial, which are associated to specific syntactic and semantic types.

Plain verbs do not express agreement morphologically. However, they use other means as in (20), in which the number is expressed in the subject, or in (21), expressed using a subject-to-object agreement auxiliary (3A.CONC.AUX.3B), also known as personal agreement marker, which is realised as a movement of the index finger from the location of the subject to the location of the object. Different values for this agreement auxiliary can be derived from the grammaticalised space shown in Fig. 4.2.

(20) 
PERSONA-PL.REP HOMBRE FÚTBOL GUSTAR
'Men like football'

(21) 
ANA PERS.3A JUAN PERS.3B 3A.CONC.AUX.3B GUSTAR
'Ann likes John'

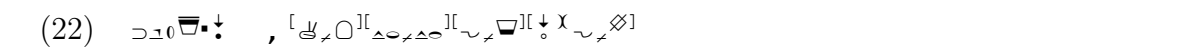
The group of plain verbs consists of impersonal verbs, unergative verbs and 'body-anchored' verbs, i.e., verbs with a specific location on the body, which represents a phonological constraint impeding the agreement with an argument. Aspect and manner can be expressed in plain verbs.

In classificatory verbs, subject agreement is expressed by substituting the configuration (or the handshape) of the verb by the configuration (or the handshape) of the subject sign or its classifier and sometimes also changing its location. Unaccusative verbs, i.e., verbs referring to visible states or processes of the subject, belong to this class of agreement verbs.

As seen in Section 2.4, there are three groups of classifiers in LSE: entity classifiers, handle classifiers and extension classifiers. Entity classifiers, which are shown in Table 2.1, are descriptive classifiers that substitute an object by its dominant dimension or shape. Even if it can be argued that classifiers can have prototypical orientations and even symmetry, this work assumes, for simplicity, that only the handshape depends on the subject and that orientation and symmetry are usually given by the verb. Handshape classifiers in Table 2.1 are incorporated by means of rewrite rules like (*4.21*), which insert the one-dimensional classifier handshape into the dominant hand in one-handed and symmetrical two-handed signs.

InsertOneDimensionalClassifier =
$(H \rightarrow$ ☝ / $\widehat{\ }S?$ ____ [?$O.^*$-CL:1D) ∘
$(H \rightarrow$ ☝ / [____ $H].^*$-CL:1D) (*4.21*)

Example (22) contains the two-handed sign 'TUMBAR (lie down)' in which the person classifier (☝) has been introflected in the dominant hand.

(22) 
PERSONA TUMBAR-CL:"PERSONA"
'A person lies down'

In example (23), the classifier of 'TELEVISIÓN (television)', which is represented by the

Orig.	Destination 1	2	3a	3b	3c	3d
1	ϵ					
2		ϵ				
3a			ϵ			
3b				ϵ		
3c					ϵ	
3d						ϵ

(a) Movements

Loc.	Destination 1	2	3a	3b	3c	3d
1						
2						
3a						
3b						
3c						
3d						

(b) Orientations

Table 4.2: (a) Movements according to its origin and destination and (b) hand orientations relative to the hand's location.

handshape configuration for framed objects (), has been introflected into the verb 'APAGAR (to switch off)'.

(23) ,
TELEVISIÓN APAGAR-CL:"TELEVISIÓN"
'Television is switched off'

Directional verbs are signs describing actions, which can be modified indicating who or what the action is being done by and to. In the case of LSE, directional verbs can be further classified into verbs with object/recipient agreement, verbs with subject and object/recipient agreement and directional-classificatory verbs. In directional verbs with object or recipient agreement, the signer assumes the role of the subject, sometimes orientating the trunk slightly towards the location it has been placed before and reproducing the face. The direction or the orientation of the sign changes to agree with the recipient or the object. Typical verbs in this group are verbs transferring attitudes, like those in the following examples:

(24) ,
PRO.1 RESPETAR-CONC.3A
'I respect her'

(25) ,
ANTONIO APOYAR-CONC.3A
'Anthony support him'

For this class of verbs, the morphological agreement tags are rewritten to a simple tags. In the previous examples, CONC.3A is rewritten as ORI.3A-DST.3A. As it has been noted in Section 4.7, HamNoSys admits the specification of absolute or relative movements. In absolute movements, only the end location of the movement is given. From this location,

the direction and size of movement is inferred. However, assuming that movements are relative, the final location must be reached from the sign's initial location following its movement. Without loss of generality, the final position of the sign is assumed to correspond to one of the locations of the grammaticalised space of Fig. 4.2. The final positions can be seen in Table 4.2.a. In order to generate the correct movements towards the final location, the initial location of the sign should be taken into consideration. This initial location can be abstracted away from its particular position considering only its position into the horizontal and vertical planes. Morpheme alternants for final position 3a are generated by rule (*4.22*), which are composed of rules as the one in (*4.23*), which deals with one specific initial location.

$$
\begin{aligned}
&\mathit{MakeMovementFinalPosition3a} = \\
&\quad \mathit{MakeMovementFrom1To3a} \circ \\
&\quad \mathit{MakeMovementFrom2To3a} \circ \\
&\quad \mathit{MakeMovementFrom3bTo3a} \circ \\
&\quad \mathit{MakeMovementFrom3cTo3a} \circ \\
&\quad \mathit{MakeMovementFrom3dTo3a} \qquad (4.22)
\end{aligned}
$$

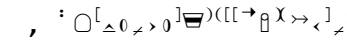
$$
\mathit{MakeMovementFrom1To3a} = (M \rightarrow \nwarrow \; / \; \blacksquare \; \_\_\_ \; \mathit{Tags}^{*}\text{-DST.3A}) \qquad (4.23)
$$

At the same time, a sign can be reoriented to its final position. This reorientation can be considered a readjust that introduces an allophonic variant, since the hand movement should be enough. Some rules for dealing with hand orientation, implemented using Table 4.2.b are:

$$
\begin{aligned}
&\mathit{MakeOrientation3a} = \\
&\quad \mathit{MakeOrientacion3aFrom1} \circ \\
&\quad \mathit{MakeOrientacion3aFrom2} \circ \\
&\quad \mathit{MakeOrientacion3aFrom3b} \circ \\
&\quad \mathit{MakeOrientacion3aFrom3c} \circ \\
&\quad \mathit{MakeOrientacion3aFrom3d} \qquad (4.24)
\end{aligned}
$$

MakeOrientation3aFrom1 = ($F \rightarrow$ / $\widehat{\ }H?^{[}?$ ___ $O.^{*}$-ORI.3A) *(4.25)*

Directional verbs with subject and object agreement use orientation to agree with its subject and use location and orientation to mark agreement with the affected object. In addition, in two-handed signs, the configuration of the passive hand sometimes agrees with the object's classifier. Typically, these verbs are transitive with a meaning of modification or translation of the object. In example (26) the subject is marked using orientation and the affected object is marked using location and orientation.

(26)  ,  ,
ANTONIO PUERTA DEM.3B / PINTAR-3B.CONC.3B
'Anthony paints that door'

Location is changed with rules like (*4.26*). Note that if the sign is two-handed with a passive hand serving as a location of the active hand, it is only the location of the passive hand that has to be relocated. In the case of two-handed symmetrical signs the change of location will affect both hands.

MakeLocation3a =
($L \rightarrow$  / $O^{]}$? ___ $^{[}?M.^{*}$-LOC.3A) ∘
($L \rightarrow$  / $^{[}L_{\swarrow}$ ___ $^{]}.^{*}$-LOC.3A) *(4.26)*

The handshape of the passive hand can be changed with rules like (*4.27*). Note that these rules are guarded to prevent other than two-handed passive-hand signs to be modified.

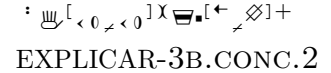
InsertOneDimensionalClassifierIntoPassiveHand =
TwoHandedPassive ∘
PrepareForTwoDifferentHandShapes ∘
($H \rightarrow$  / $^{[}H_{\swarrow}$ ___ $^{]}$) *(4.27)*

When both hands have the same handshape, rule (*4.28*) prepares the sign for having a different handshape in each hand inserting the appropriate symbols.

$$PrepareForTwoDifferentHandShapes =$$
$$(\epsilon \rightarrow {}^{[} \ / \ L \ \_\_\_ \ H) \circ (\epsilon \rightarrow {}_{\nearrow}H^{]} \ / \ L^{[}H \ \_\_\_ \ ) \qquad (4.28)$$

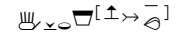
For directional verbs with subject and recipient agreement, the subject agreement applies to its location and orientation and the recipient agrees with the orientation and the direction of the movement, which ends up at the recipient's location. In general, this subclass of verbs correspond generally with ditransitive verbs transferring information, as 'EXPLICAR (to explain)', and with deontic verbs, as 'AUTORIZAR (to authorise)' or 'PROHIBIR (to forbid)'. Note that the verb 'MIRAR (to look)' belongs to this class since, for the Deaf, it is an information providing action. Examples in (27) and (28) mark the subject using location and orientation and the recipient using orientation and the direction of movement. In example (27), 3B.CONC.3A is replaced by LOC.3B-ORI.3A-DST.3A.

(27) 
PREGUNTAR-3B.CONC.3A
'He ask him'

(28) 
EXPLICAR-3B.CONC.2
'He explains you'

Some verbs are called reversible when the interchanging subject and recipient do not imply changing location, but only its orientation and the direction of movement. This is the case of 'COMPRENDER (to understand)', as it can be seen in the following examples:

(29) 
COMPRENDER-1.CONC.2
'I understand you'

(30) 
COMPRENDER-2.CONC.1
'You understand me'

Verbs with a reciprocal meaning become two-handed symmetrical to represent the two subjects, thus expressing a form of plural, as in the following examples:

(31) 
MIRAR-3A.CONC.RECP.3B
'They look each other'

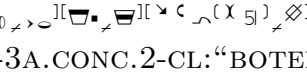
(32) 
ESTAR.DE.ACUERDO-3A.CONC.RECP.3B
'They both agree'

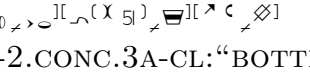
To make one-handed signs symmetrical is not trivial in HamNoSys. In fact, the previous examples are relatively simple since they are reflections on the horizontal plane and there exists a HamNoSys operator for this kind of symmetry. However, location, orientation and movement of the non-dominant hand has to be inferred from the parameters of the dominant. In the case of 3A.CONC.RECP.3B, the tag is replaced by 3A.CONC.3B and 3B.CONC.ND.3A, which in turn are replaced by LOC.3A-ORI.3A-DST.3B and LOC.ND.3B-ORI.ND.3B-DST.ND.3A respectively. Note that tags containing ND affect the non-dominant hand. A set of rewrite rules like *MakeNonDominantHandLocation3a*, *MakeNonDominantHandOrientation3a* and *MakeNonDominantHandFinalPosition3a* mimicking the operations the dominant hand has to be created.

Directional-classificatory verbs are ditransitive verbs that agree with the subject, the object and the recipient or location. Object agrees through its classifier. Initial location marks the deficitary, which can be the subject, as in 'DAR (to give)', the recipient, as in 'ROBAR (to steal)', the locative, as in 'PONER (to place)'. Movement is directed to the beneficiary, which can be the subject, as in ROBAR, the recipient, as in DAR, or the location, as in PONER. This can be seen in the following examples:

(33) 
LIBRO ROBAR-3A.CONC.3B-CL:"LIBRO"
'He steals him the book'

(34) 
LIBRO DAR-3A.CONC.3B-CL:2D
'He gives the book to him'

(35) 
MESA DEM.2 / BOTELLA

PONER-3A.CONC.2-CL:"BOTELLA"-CLP:"MESA"
'Put the bottle on the table'

(36) 
MESA DEM.2 / BOTELLA

COGER-2.CONC.3A-CL:"BOTTLE"-CLP:"TABLE"
'Take the bottle from the table'

Classifiers are inserted using previously defined rules as *InsertOneDimensionalClassifier* and the rules used by other verb classes to locate signs and to determine its directional

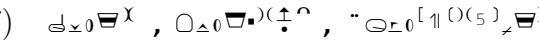
movements. According to the starting of the action, verbs can be classified as regular or backward. Regular verbs start its movement at the subject, as do DAR. On the contrary, backward verbs start their movements at the object, as the sign 'COGER (to take)' does.

Spatial verbs have descriptive movements, which are not grammatical. Apart from the spatial information they transmit, these verbs can also convey information on the action's participants by means of classifiers. We will not address this kind of verbs here, but some approaches can be found in Huenerfauth (2004).

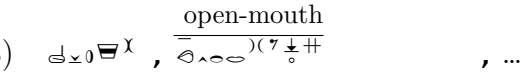
4.10 Aspectual Marking

The category of aspect describes the internal time of events and states. Generally speaking, languages distinguish two basic aspect values: perfective, when the verbal action is perceived as finished, and imperfective, when the verbal action is in development or in process. In LSE, predicative verbs and adjectives display aspect values. These values are normally expressed in LSE using lexical items, but sometimes they are expressed morphologically using the movement and NMCs. Aspect values that can be expressed morphologically are: inchoative, punctual, perfective or terminative, imperfective or durative, iterative, frequentative, gradual and progressive.

Inchoative aspect value indicates that the action or state is going to start. In LSE, this aspectual value can be expressed in several ways, by adding to the sign the verb 'PREPARAR (to prepare)', the adverb 'CASI (almost)', or using a periphrastic construction with the verb 'IR (to go)', as in example (37).

(37) 
PERS.1 IR TRABAJAR
'I'm going to work'

Morphologically, inchoative aspect is also expressed suspending for a while the execution of the sign and an open mouth, as in example (38).

(38) 

PERS.1 COMER-ASP.INCHOAT …
'I was going to eat …'

If the suspension is caused by an impossibility, the tight-lipped is used instead of an open mouth. Both forms of expression using NMCs give rise to different rewrite rules:

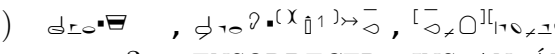
$$InchoativeAspect = InsertHoldAtStart \circ InsertOpenMouth \tag{4.29}$$

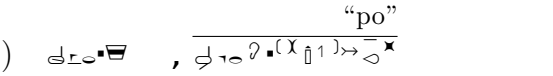
$$InchoativeAspectWithImpossibility = InsertHoldAtStart \circ InsertTightLips \quad (4.30)$$

$$InsertHoldAtStart = (\epsilon \rightarrow \text{ʾ} \ / \ L \ \_\_\_ \ M) \mid ((\epsilon \rightarrow \text{ʾ} \ / \ [\ \_\_\_ \ M) \circ (\epsilon \rightarrow \text{ʾ} \ / \ \text{ˏ} \ \_\_\_ \ M \backslash \not{\text{ˏ}})) \quad (4.31)$$

In addition to the use of NMCs, a pause with a second of duration can be introduced in the middle of the performance or the sign. Even the second half of the sign can be suppressed after the pause. However these forms of realisation are more difficult to implement since the movement should be segmented. We do not provide rules for dealing with these implementations of the inchoative aspect value.

Punctual aspect indicates that the action is instantaneous and can be expressed adding the sign 'INSTANTÁNEO (instantaneous)' after the verb, as in example (39), or with NMCs, as the abbreviation or intensification of the movement, the phonation "po", etc., as in example (40).

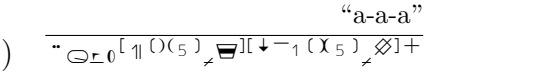
(39) 
PERS.3A ENSORDECER INSTANTÁNEO
'He deafened'

(40) 
PERS.3A ENSORDECER-ASP.PUNCT
'He deafened'

In the implementation of punctual aspect marking, the intensification of movement is assimilated to tension, realised by the rule *MakeMovementIntense*, which is implemented similar to the rule *InsertHoldAtStart* in (*4.31*), but introducing the symbol '✖' for tension.

$$PunctualAspect = MakeMovementIntense \circ InsertMouthing\text{-}\text{“}po\text{”} \quad (4.32)$$

Imperfective or durative aspect indicates that the action is still in development without interruption and is usually expressed in LSE with a repetition of the sign. Also, some NMCs are usually added to the repetition, like a slight vibration of lips, the mouthing "a-a-a", as in the example (41), or "ze-ze", or a movement slowdown.

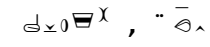
(41) 
TRABAJAR-ASP.DUR

'To be working'

An implementation of this aspect value can be seen in rule (*4.33*). The rule *InsertInPlaceRepetition* was developed in Section 4.7 for plurals.

$$ImperfectiveAspect = InsertInPlaceRepetition \circ InsertMouthing\text{-}\text{“}a\text{-}a\text{-}a\text{”} \quad (4.33)$$

Perfective or terminative aspect, indicating that the action is finished, is expressed preferably by adding the adverb 'YA (already)' after the verb, as in the following example:

(42) 
PERS.1 ARREGLAR YA
'I fixed'

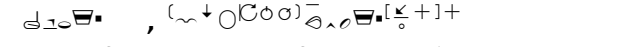
Alternatively, this aspect value can be expressed using NMCs such as the suspension of the sign at the end of the movement and the mouthing 'apr', a coarse tongue, etc. Assuming that suspension can be implemented as a sudden stop at end, perfective aspect is realised with the following rules:

$$PerfectiveAspect = InsertSuspension \circ InsertMouthing\text{-}\text{“}apr\text{”} \circ InsertCoarseTongue \quad (4.34)$$

$$InsertSuspension = (\epsilon \rightarrow \shortparallel \ / \ M \ \_\_\_ \ Tags) \quad (4.35)$$

For the iterative aspect, 'OTRA.VEZ (again)' can be added after the verb, as in example (43), or the sign can be repeated, with the specific NMCs like the semi-open mouth, frowned eyebrows, head bowed, etc., as in example (44).

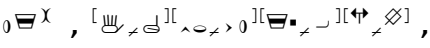
(43) 
PERS.3B LLAMAR-3A.CONC.1 OTRA.VEZ
'He calls me repeatedly'

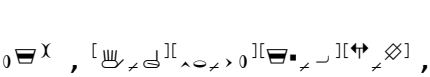
(44) 
PERS.3B LLAMAR-3A.CONC.1-ASP.ITER
'He calls me repeatedly'

The *IterativeAspect* rule is implemented as follows:

$$IterativeAspect = InsertInPlaceRepetition \circ InsertMouthing\text{-}\text{“}apr\text{”} \circ InsertCoarseTongue \quad (4.36)$$

Frequentative aspect, indicating that the action is habitual, is marked lexically with the sign 'FRECUENTEMENTE (frequently)' after the verb, as in example (45), or it can be expressed morphologically, by means of repetition together with the mouthing 'a-a-a', as in example (46):

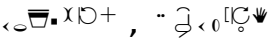
(45) 
PERS.1 CINE IR FRECUENTEMENTE
'I usually go to the cinema'

(46) 
PERS.1 CINE IR-ASP.FREQ
'I usually go to the cinema'

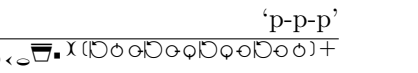
This aspect value is implemented as follows:

$$FrequentativeAspect = InsertInPlaceRepetition \circ InsertMouthing\text{-}\text{“a-a-a”} \quad (4.37)$$

Finally, gradual aspect is expressed with the adverb 'POCO.A.POCO (little by little)' after the verb, as in example (47).

(47) 
PRACTICAR POCO.A.POCO
'Practicing'

However, it can also be expressed by NMCs, with the mouthing 'p-p-p' and making the movement staccato, as in example (48).

(48) 
PRACTICAR-ASP.GRAD
'Practicing'

Note that there is no easy way to make a movement staccato in HamNoSys since movement should first be parsed in order to obtain its sequential segments. However, a HamNoSys modality value for staccato could be convenient as there is a value for tense movements and also considering its ease of implementation for its synthesis with avatar technology.

4.11 Adjectives

Nouns usually precede adjectives, but some adjectives can be modified by introflection to introduce the configuration of the noun or the configuration of the classifier of the

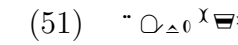
noun. In the absence of articulatory constraints, the possibility of introflection depends on the semantics of the noun and the adjective. Concrete nouns, or their classifiers, can be introduced into adjectives assigning physical properties. Incorporation of the noun configuration is found in size adjectives, like 'ALTO (tall)' / 'BAJO (short)' and 'GRANDE (big)' / 'PEQUEÑO (small)', and in shape adjectives, like 'LARGO (long)' / 'CORTO (short)', 'ESTRECHO (narrow)' / 'ÁNCHO (wide)' and 'CUADRADO (squared)' / 'CIRCULAR (circular)'. In example (49), the configuration of 'CUADRO (picture)' is incorporated to the adjective 'GRANDE (big)'.

(49)  ,
CUADRO GRANDE-G:"CUADRO"
'a big picture'

Classification is found in adjectives of posture, like 'TORCIDO (crooked)' or 'APILADO (stacked)', and in adjectives of animation, like 'VOLADOR (flying)' or 'RÁPIDO (fast)' / 'LENTO (slow)'. The classifier of the sign 'ESPEJO (mirror)' is introflected into the adjective TORCIDO in (50).

(50)  ,
ESPEJO TORCIDO-CL:"MIRROR"
'a crooked mirror'

However, adjectives like 'GRUESO (wide)' have different morphological realisations. In example (51), GRUESO is realised with a classifier, but in example (52), it incorporates the configuration of the noun.

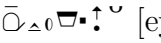
(51)  ,
LIBRO GRUESO-CL:"LIBRO"
'a wide book'

(52)  ,
VASO GRUESO-G:"GLASS"
'a wide glass'

Introflection is usually accompanied by NMCs, realised as changes in the movement and in its speed, and by different degrees of articulatory tension, especially for consistency and tactile feeling adjectives: 'DURO (hard)' / 'BLANDO (soft)', 'CONSISTENTE (consistent)' / 'FRÁGIL (fragile)', 'PESADO (heavy)' / 'LIGERO (light)' and 'RUGOSO (rough)' / 'SUAVE (smooth)'.

4.12 Degree

As for adverbs and verbs, adjectives admit the degree values of intensification and restriction, realised as NMCs, which are equivalent to adverbs. Apart from properties with specific realisations of intensification or restrictions, if the property is positive, the intensification is performed with clenched teeth, half-open eyes and articulatory tension, as in example (53).

(53)  [eye_lids=SB] [mouth_gesture=D01]
ALTO-INTENS
'very high'

Negative properties are intensified with inflated cheeks, blowing and articulatory tension, as in example (54).

(54)  [mouth_gesture=L20]
PEQUEÑO-INTENS
'very little'

Restriction is expressed by arching lips and a lateral inclination of the head, as in the following example:

(55)  [mouth_gesture=L24] [head_movement=TL]
ALTO-RESTR
'not very high'

Additionally, some adjectives show aspect values, which are realised in the same way it is described for verbs in Section 4.10.

4.13 Implementation

As a proof of concept, the morphological grammar described in the above sections has been compiled to transducers using the OpenGrm Thrax Grammar Compiler (Tai, Skut and Sproat, 2011). HamNoSys symbols have been taken from SiGML (Signing Gesture Markup Language), an XML application based on HamNoSys that provides HamNoSys symbol names for the encoding of manual features. SiGML is the input notation used by the JASigning application, a synthetic sign language performance system superseding the earlier software developed in ViSiCAST (Schulmeister, 2001) and eSIGN projects (Kennaway, Glauert and Zwitserlood, 2007). A symbol as '' is defined as a Thrax symbol using its SiGML name '`[extfingeru]`':

```
extfingeru = "[extfingeru]";
```

The previous and other symbol definitions are then used in other definitions, as in the following example, where the orientation parameter O is defined:

```
O = finger_orientation palm_orientation ;
palm_orientation = palmu | palmur | ... | palmul ;
finger_orientation = finger_base_orientation finger_base_orientation? ;
finger_base_orientation = extfingeru | extfingerur | ... | extfingeruo ;
```

An example of how rewrite rules are implemented using Thrax can be seen in the following fragment of code, which corresponds to rules described in Section 4.7:

```
RemoveTargetLocation =
  CDRewrite[T : "", parbegin R, plus M parend, Sigma*] @
  CDRewrite[T : "", parend? R, parbegin? SigmaTag* "[EOS]", Sigma*] ;

ProjectDominantHand =
  CDRewrite[S : "", "", "", Sigma*] @
  CDRewrite[H|O|L|M : "", plus, parend, Sigma*] @
  CDRewrite[parbegin : "", "", (H|O|L|M) plus parend, Sigma*] @
  CDRewrite[plus parend : "", "", "", Sigma*] ;

InsertParallelSymmetry =
  CDRewrite["" : symmpar, "[BOS]", "", Sigma*] ;

InsertAlternatingMovement =
  CDRewrite["<" : seqbegin, "", "", Sigma*] @
  CDRewrite[">" : seqend alternatingmotion, "", "", Sigma*] @
  CDRewrite["<"|">" : "", "", "", Sigma*] ;

RepeatAlternatingHandsHorizontally =
  RemoveTargetLocation @
  ProjectDominantHand @
  InsertParallelSymmetry @
  MarkMovement @
  InsertAlternatingMovement ;
```

For the non-manual component, SiGML includes different tiers for shoulders, body, head, eye-gaze, facial expression and two sub-tiers for mouthing. SAMPA codes are used for speech mouthing.

Unfortunately, the implementation of the morphological model presented here is difficult to evaluate providing that LSE, and sign languages in general, are under-resourced languages. To our knowledge, there are no HamNoSys texts in LSE that could be analysed with a lexicon to measure its coverage, nor another morphological generator to measure its relative precision and recall.

Chapter 5

Conclusions and Future Work

The choice of a particular type of technology to process a language or a variety of a language is greatly influenced by the density of the language, i.e., the availability of digital stored resources. Commercial research and development have concentrated on high-density languages. According to Varga et al. (2007), for medium-density languages, which are spoken by over half of humanity, parallel corpora can be compiled using digital resources like literary and religious texts, international law, movie captioning, software documentation, bilingual magazines, corporate home pages and annual reports. Today, Spanish Sign Language (LSE), like any other sign language, is a low-density or under-resourced language. Because of modality, acquisition of sign language data is a time-consuming and expensive task compared to the acquisition of spoken or written data. For LSE, no parallel corpus currently exists of sufficient size to enable data-driven approaches to machine translation in non-restricted domains. However, LSE is being standardised and the language's first grammar, as well as a normative dictionary, have been published a few years ago. All these factors make the rule-based approach the only viable approach today for the development of a production-quality Spanish-to-LSE machine translation system for non-restricted domains.

This thesis has presented a transfer-based approach for Spanish-to-LSE translation by delivering LSE phonological representations. A wide-coverage Spanish parser is used to obtain a dependency analysis. The part-of-speech flexibility of LSE enables the augmentation of the bilingual lexicon by bootstrapping the initial lexicon using morpho-lexical relationships. Lexical-semantic relationships are used to bridge the lexical gap and classifier names are also generated using lexical-semantic relationships. A word order generation algorithm has been presented to deal with the topic-oriented surface order of LSE. The algorithm make use of linear precedence rules operating at the level of syntactic functions and dealing with topicalisation. Produced glosses are annotated with number, aspect, or other morphological information. A computational morphology for a significant fragment

the of LSE inflectional morphology has been implemented. It takes annotated glosses and produces HamNoSys (Hamburg Sign Language Notation System) phonological representations. Finally, classifier predicates, which are one of the cornerstones of sign languages, has a translatable subclass corresponding to Spanish prepositional expressions with locative and temporal meanings. Moreover, some of the solutions presented in this thesis to the problems found for machine translation between Spanish and LSE could be equally valid, with some adaptations, for the translation between other spoken languages and other sign languages.

A parallel Spanish-LSE corpus has been created from the data used in a psycholinguistic study. Although the corpus is comparable in size to other corpora used in data-driven approaches, it is not domain-specific. A parallel testbed has been used to evaluate the system. Experiments reported a BLEU (Bilingual Evaluation Understudy) about 0.30 and a TER (Translation Error Rate) at about 42%. On the one hand, these results confirm the intuition that when data are scarce, or in sparse domains, induction is difficult or impossible, and hence bilingual data and handcrafted grammars are good alternatives. On the other hand, experiments show that LSE word order generation has played some little role in this corpus. This does not mean that word order in LSE and in Spanish is the same. In fact, they have been reported to be very different. However, as seen in the backtranslations of the corpus, the order in Spanish is free enough to accommodate the sign ordering of the LSE sentences. Finally, a linguistic-oriented error analysis has shown that many differences between the system output and the reference translations arise from variations in the realisation of the linguistic structures and that classifier predicates are the most difficult expressions to generate.

Despite the relatively good results obtained with the approach presented in this thesis, experiments should be extended in several directions in order to assess properly the performance of the system. The corpus should incorporate more linguistic phenomena in order to evaluate the coverage of the system and its components. In addition, the corpus should conform to standard LSE, since, from a sociolinguistic point of view, the standard is the variant used in addressing a linguistic community, in this case the LSE community. This thesis is also incomplete, since it addresses only the translation of written Spanish-to-LSE glosses and then into phonological representations. Glosses are used as an intermediate symbolic representation for guiding the development and assessing the performance of the system in sign choices and in generating sign sequences. However, glosses are merely an underspecified symbolic representation of a signed message. Phonological representations are not even a writing system for sign languages that could be understood by Deaf people. A complete machine translation system for sign languages should produce animations and a genuine and proper evaluation should involve Deaf people and should

measure comprehension. Many other important aspects have been left unaddressed for the moment and there is still a great deal of work to do, even in the field of animation synthesis, in order to make machine translation systems operational and useful.

Appendix A

Corpus

(1) a. Con el mes de septiembre para muchos estudiantes comienza una nueva y desafiante etapa en su vida: la entrada en la Universidad.

b. TOCA MES SEPTIEMBRE PERSONA+ ESTUDIANTE+ EMPEZAR ÉPOCA ESPECIAL VIDA SU, ENTRAR UNIVERSIDAD.

c. En el mes de septiembre los estudiantes empiezan una época especial de su vida, la entrada en la universidad.

(2) a. Casi la mitad de los estudiantes de la Universidad de Sevilla viene de los pueblos de la provincia y de otras ciudades andaluzas.

b. CASI MITAD PERSONA+ ESTUDIANTE+ UNIVERSIDAD SEVILLA VENIR DÓNDE-?, PUEBLO SU, PROVINCIA SEVILLA O PROVINCIA OTRO.

c. Casi la mitad de los estudiantes de la Universidad de Sevilla vienen de los pueblos de su provincia o de otras provincias.

(3) a. El primer y principal problema con que se encuentran estos alumnos es la búsqueda de una vivienda.

b. PRIMERO PROBLEMA CUÁL-?, PERSONA+ ESTUDIANTE+, PROBLEMA CUÁL-?, BUSCAR PISO, CASA, VIVIR DÓNDE-?.

c. El primer problema de los estudiantes es buscar un piso, una casa donde vivir.

(4) a. ¿Dónde vivir? es la pregunta que a muchos estudiantes no deja dormir alguna que otra noche.

b. VIVIR DÓNDE-? MUCHAS-VECES POR-LA-NOCHE CLS: 2 "persona dando vueltas en la cama" PREOCUPAR+ PROBLEMA ESE.

c. ¿Dónde vivir? Muchas veces por la noche no pueden dormir preocupados por ese problema.

(5) a. Los que tienen más suerte son aquellos que tienen hermanos o amigos que ya viven en Sevilla y que les ayudan a buscar piso.

b. ALGUNOS SUERTE HAY AMIGO O HERMANO, ETCÉTERA, ESOS PERSONA++ APOYARME PARA BUSCAR PISO.

c. Algunos tienen suerte y tienen amigos o hermanos que les apoyan para buscar piso.

(6) a. Pero son muchos los que llegan perdidos y asustados y no saben por dónde empezar a buscar.

b. PERO, MUCHAS-VECES ALGUNOS ABRIR-PUERTA NO-SABER IR DÓNDE MOTIVO BUSCAR ZONA NO-SABER DÓNDE.

c. Pero, muchas veces, algunos no saben por dónde empezar porque no saben buscar la zona.

(7) a. Se les encuentra por todas partes buscando un piso donde vivir el resto del año.

b. MUCHAS-VECES SEVILLA ZONA BUSCAR+ CLP: HAY HAY HAY "se signa en distintos espacios" PARA VIVIR DURANTE-UN-AÑO.

c. Muchas veces buscan por Sevilla donde vivir durante el año.

(8) a. A la hora de plantearse vivir fuera de casa, son muchas las alternativas a las que podemos acogernos.

b. TOCA EMPEZAR CASA VIVIR DENTRO-NO, FUERA HAY POSIBILIDAD CLP: 4 4 "muchas posibilidades" (mofl),

c. Cuando empiezan a vivir fuera de casa hay muchas posibilidades.

(9) a. La más extendida quizá entre los estudiantes, es la de alquilar un piso entre varios amigos.

b. MAYORÍA USAR+ PERSONA+ ESTUDIANTE+ USAR CUÁL-?, ALQUILER+ AMIGO COMPARTIR AMIGO.

c. La mayoría usa el alquiler compartido con amigos.

(10) a. Esta es la solución que prefieren los estudiantes.

b. ESE, ALQUILER PISO GUSTAR MÁS QUIÉN, PERSONA+ ESTUDIANTE+.

c. El alquiler de piso le gusta más a los estudiantes.

(11) a. Esta opción tiene como ventajas las siguientes:

b. ESE ALQUILER POSITIVO++ QUÉ-?:

c. El alquiler tiene cosas positivas:

(12) a. si se alquila el piso entre varios amigos,

b. EJEMPLO, PRIMERO AMIGO+ ALQUILER DINERO COMPARTIR,

c. Por ejemplo, primero, el dinero del alquiler se comparte con los amigos

(13) a. el alojamiento sale más barato.

b. REBAJAR DINERO REBAJAR.

c. el precio baja.

(14) a. los estudiantes tienen libertad para entrar y salir de su casa cuando quieran.

b. SEGUNDO, PERSONA ESTUDIANTE+ LIBRE PODER ENTRAR, VENIR, COMOQUIERA, LIBRE.

c. Segundo, los estudiantes tienen libertad para poder entrar y venir cuando quieran.

(15) a. los pisos de alquiler suelen estar cerca de las distintas facultades. Por lo tanto, se ahorran el dinero de tener que coger el autobús.

b. TERCERO, QUÉ-?, ALQUILER+ CERCA UNIVERSIDAD MOTIVO TEMA DINERO+ AUTOBÚS ECONOMÍA AHORRAR++.

c. Tercero, el alquiler cerca de la universidad para ahorrar el dinero del autobús.

(16) a. Por otra parte, hay que tener en cuenta que alquilar un piso también tiene algunos inconvenientes.

b. CUARTO, TAMBIÉN, ALQUILER PISO, TAMBIÉN COSAS NEGATIVO++ HAY, CUÁL-?,

c. En cuarto lugar, el alquiler también tiene cosas negativas.

(17) a. Por ejemplo, el estudiante tiene que hacerse la comida él sólo, lavarse y plancharse su ropa y limpiar el piso.

b. EJEMPLO, PERSONA ESTUDIANTE+ DEBER COMIDA A-MANO COCINAR+ SOLO COCINAR++, ROPA LAVADORA O FREGAR, ETCÉTERA.

c. Por ejemplo, los estudiantes deben hacerse su propia comida, deben cocinarse solos, lavar la ropa o fregar, etcétera.

(18) a. Para encontrar piso, sólo basta darse un paseo por la facultad e ir bien atento mirando las paredes,

b. TOCA ENCONTRAR PISO SÓLO HACER-FALTA UNIVERSIDAD CARTEL CLL: "cartel en la pared" CLP: 5 "muchos carteles en las paredes" VER,

c. Para encontrar piso sólo hacen falta los carteles de la universidad.

(19) a. porque, cuando llega septiembre, éstas se llenan de anuncios de pisos de distintos precios.

b. TOCA SEPTIEMBRE, PARED ESPECIAL UNIVERSIDAD ANUNCIO DINERO ETCÉTERA.

c. Cuando llega septiembre en las paredes de la universidad hay anuncios con el precio, etcétera.

(20) a. Los precios de los alquileres varían dependiendo de la zona en la que esté el piso.

b. PRECIO ALQUILER DEPENDER (2m)QUÉ-?, ZONA DEPENDER MOTIVO ALQUILER+ DINERO VARIAR++.

c. El precio del alquiler depende de la zona porque el dinero de los alquileres varía.

(21) a. Una solución distinta al piso de alquiler es la residencia universitaria.

b. OTRO SOLUCIÓN CUÁL-?, RESIDENCIA ESPECIAL UNIVERSIDAD.

c. Otra solución es la residencia universitaria.

(22) a. Esta es la solución que prefieren los padres de los estudiantes.

b. ESE RESIDENCIA GUSTAR QUIÉN-?, PADRE-MADRE PERSONA ESTUDIANTE+

c. La residencia le gusta a los padres de los estudiantes

(23) a. Los padres se sienten más tranquilos y relajados con la residencia

b. MOTIVO PADRE-MADRE TRANQUILO, RELAJADO TEMA RESIDENCIA

c. porque se sienten tranquilos y relajados con la residencia

(24) a. porque piensan que aquí los estudiantes están más controlados y, además, reciben educación y una buena y sana alimentación.

b. MOTIVO PADRE-MADRE PENSAR DENTRO PERSONA ESTUDIANTE+ CONTROL, MÁS EDUCACIÓN ESPECIAL, ADEMÁS COMIDA CALIDAD BUENO.

c. porque los padres piensan que dentro los estudiantes están controlados, tienen una educación especial y buena calidad en las comidas.

(25) a. Éstas son algunas de las ventajas de las residencias universitarias, pero otras ventajas son:

b. ESE ANTES GUIÓN++ POSITIVO++, OTRO RESIDENCIA POSITIVO++ (2m)QUÉ-?:

c. Esto son cosas positivas, otras cosas positivas de las residencias son:

(26) a. existen muchas residencias en Sevilla y de distinto tipo, por ejemplo, las hay femeninas, masculinas y mixtas.

b. DENTRO SEVILLA CIUDAD RESIDENCIA ETCÉTERA, EJEMPLO, HAY MUJER GRUPO O HOMBRE O MEZCLADO.

c. Dentro de la ciudad de Sevilla hay distintas residencias, por ejemplo, femeninas, masculinas y mixtas.

(27) a. hay residencias por todas partes

b. SEGUNDO, RESIDENCIA CLP: "muchas residencias por todas partes" CIUDAD.

c. Segundo, hay residencias en todas partes de la ciudad.

(28) a. y los estudiantes pueden elegir aquella que esté más cerca de su facultad.

b. ADEMÁS ESTUDIANTE+ PODER IR CERCA UNIVERSIDAD IR.

c. Además, los estudiantes pueden ir a la que está cerca de la universidad.

(29) a. en las residencias se hacen muchos amigos y se conoce a mucha gente.

b. DENTRO RESIDENCIA AMIGO ETCÉTERA PODER, CONOCERSE+++.

c. Dentro de la residencia se pueden conocer amigos.

(30) a. algunas residencias dan clases de apoyo para que el estudiante vaya bien en sus estudios.

b. ALGUNOS RESIDENCIA HAY DENTRO CLASE+ APOYO PARA PERSONA ESTUDIANTE+ MEJOR+, ESTUDIO MEJOR++.

c. Algunas residencias tienen clases de apoyo para que los estudiantes vayan mejor en sus estudios.

(31) a. Pero la residencia universitaria también tiene inconvenientes.

b. DENTRO RESIDENCIA COSAS NEGATIVO+,

c. Dentro de la residencia hay cosas negativas,

(32) a. Por ejemplo, los estudiantes se siente vigilados porque la hora de cierre de la residencia suele ser muy temprana y cada vez que se marchan de fin de semana deben rellenar un parte de salida.

b. EJEMPLO, PERSONA ESTUDIANTE SENTIR CONTROL MOTIVO HORA CERRAR PUERTA TEMPRANO PROBLEMA SALIR NO-PODER, ADEMÁS FIN-DE-SEMANA IR CASA, DEBER JUSTIFICANTE FIRMAR PRESENTAR.

c. por ejemplo, los estudiantes se sienten controlados porque la hora de cierre de la puerta es temprano y tiene el problema de no poder salir. Además, el fin de semana se van a casa y deben presentar un justificante firmado.

(33) a. La residencia universitaria es más cara que el piso de alquiler.

b. APARTE RESIDENCIA UNIVERSIDAD MÁS-CARO DINERO MÁS-CARO, PISO MENOS.

c. Por otra parte, la residencia universitaria es más cara que un piso.

(34) a. La guerra es tan vieja como la humanidad.

b. GUERRA EDAD IGUAL HUMANIDAD IGUAL+.

c. La guerra tiene la misma edad que la humanidad.

(35) a. Entre los hallazgos más antiguos están las armas de guerra: hachas, lanzas, flechas, espadas y puñales en épocas prehistóricas.

b. COSAS ENCONTRAR, COSAS ANTIGUO (2m)QUÉ-?, GUERRA ESPECIAL GUERRA COSAS, LANZA, PALO, FLECHA, ESPADA, PUÑAL, ENCONTRAR DÓNDE-?, ÉPOCA ANTES HISTORIA, ANTES.

c. Se han encontrado antigüedades especiales para la guerra. Lanzas, palos, flechas, espadas, puñales de la prehistória.

(36) a. Hay diferencias en la agresividad de los pueblos,

b. TEMA GRUPO++ AGRESIVIDAD++ ESE DIFERENTE++.

c. La agresividad es diferente en los grupos.

(37) a. por ejemplo, no existen armas de guerra y sí sólo de caza entre los esquimales, ciertas tribus de Colombia, Ecuador y África.

b. EJEMPLO, GRUPO ALGUNOS HAY GUERRA NO-HAY, MATAR ANIMALES SÍ, GRUPO QUIÉN-?, EJEMPLO VIVIR PERSONA+ POLO NORTE ESQUIMAL, GRUPO COLOMBIA, ECUADOR, ÁFRICA.

c. Por ejemplo, algunos grupos no tienen guerra pero sí matan animales. Esos grupos son, por ejemplo, los esquimales del Polo Norte, grupos de Colombia, de Ecuador y de África.

(38) a. En sus lejanos espacios vitales, poco poblados y de difícil acceso, no había enemigos.

b. ZONA SU LEJOS, PERSONA+ DENTRO VIVIR POCO, PODER IR COSTAR, PERSONA ENEMIGO NO-HAY.

c. En sus zonas lejanas vivían pocas personas, costaba poder llegar, no tenían enemigos.

(39) a. Existen en cambio un sinfín de ejemplos de pueblos agresivos, sanguinarios y belicosos:

b. ENTONCES, PERO HAY EJEMPLO GRUPO AGRESIVIDAD+++, GANAS SANGRE, AMOR GUERRA,

c. Pero hay ejemplos de pueblos muy agresivos, sanguinarios y amantes de la guerra.

(40) a. destacaron, entre otros, por su sed de sangre y afán destructor, los Zulúes de África del Sur.

b. PRIMERO IMPORTANTE QUIÉN-? MOTIVO GANAS MATANZA GANAS QUIÉN-?, GRUPO NOMBRE DL-ZULÚES. DÓNDE VIVIR-?, ÁFRICA SUR.

c. El más importante por sus ganas de matanza era el grupo llamado Zulúes, que vivían en África del Sur.

(41) a. Desde tiempos muy remotos fueron causa de guerra el robo de mujeres y tierras, la sed de botín, las ideas religiosas,

b. TOCA HACE-MUCHO-TIEMPO++ OBJETIVO GUERRA (2m)QUÉ-?, MUJER (2m)ROBAR, TIERRA (2m)ROBAR+, TAMBIÉN GANAS BOTÍN, IDEA RELIGIÓN,

c. Hace mucho tiempo el objetivo de la guerra era el robo de mujeres y tierras, también las ganas de botín, las ideas religiosas,

(42) a. que subyacen a costumbres como los sacrificios humanos, el canibalismo o la caza de cabezas y no en último lugar el ansia de gloria y la voluntad de dominio.

b. DENTRO RELIGIÓN HAY COSTUMBRE EJEMPLO PERSONA MUERTO+++ MATANZA O EJEMPLO PERSONA CUERPO COMER CLC: "comerse el cuerpo" O PERSONA CABEZA CORTAR-CABEZA, TAMBIÉN, ÚLTIMO-NO, QUÉ-?, GANAS ÉXITO, MÁS GANAS DOMINAR CONTROL.

c. dentro de la religión había la costumbre de, por ejemplo, matar a las personas o, por ejemplo, comer el cuerpo de las personas o cortarles la cabeza. También, no en último lugar, estaban las ganas de éxito, las ganas de dominar y controlar.

(43) a. Los imperios de la historia se formaron basándose en guerras.

b. GRUPO GANAS DOMINAR, NOMBRE DL-IMPERIOS, EJEMPLO GRECIA ROMA, BASE SURGIR MOTIVO GUERRA.

c. El grupo con ganas de dominar recibía el nombre de imperios. Por ejemplo, Grecia, Roma, surgieron basándose en las guerras.

(44) a. En casi todas las tribus y pueblos gozó el guerrero de especial prestigio.

b. CASI MAYORÍA GRUPO, PUEBLO, IMPORTANTE QUIÉN-?, HOMBRE PERSONA RESPONSABLE GUERRA,

c. Casi en la mayoría de los grupos, pueblos, el importante era el guerrero.

(45) a. Su posición en la sociedad era elevada, se vestía con espléndido, magnífico y llamativo atuendo para la lucha

b. DENTRO SOCIEDAD NIVEL ALTO, ROPA MARAVILLOSO, RESTO LLAMAR-ATENCIÓN, ESPECIAL PARA LUCHAR.

c. Dentro de la sociedad tenía un nivel alto, una ropa maravillosa que llamaba la atención de los demás y era especial para la lucha.

(46) a. a la que solía preceder un ritual destinado a levantar el ánimo, la danza guerrera.

b. ESE ANTES LUCHAR ANTES HAY SABER BAILE ESPECIAL PARA GUERRA NOMBRE PARA ANIMAR+.

c. Antes de la lucha hay un baile especial para la guerra llamado para animar.

(47) a. Hubo siempre gran interés en una presentación aterradora, amenazante y hostil, para ello se usaban pinturas de guerra y máscaras.

b. SIEMPRE++ PRESENTACIÓN EJEMPLO MIEDO, PODEROSO, MÁS PARA RESTO (2m)ASUSTAR, ENTONCES MUCHAS-VECES USAR PINTURAS COLOR ETCÉTERA, TAMBIÉN MÁSCARA.

c. Siempre, siempre, siempre se presentaban dando miedo, con poder, para asustar al resto, entonces, muchas veces usaban pinturas de colores variados y máscara.

(48) a. En sus comienzos la guerra consistía en una lucha mano a mano entre tribus contrarias tras el encuentro inicial de los campeones.

b. AL-PRINCIPIO TEMA GUERRA ENFRENTARSE GRUPO ENEMIGO, PRIMERO PERSONA+ CAMPEÓN+ ENCONTRARSE, DESPUÉS GRUPO CLS: 4 4 “dos grupos enfrentándose”.

c. Al principio, en las guerras, se enfrentaban los grupos enemigos y primero se encontraban los campeones y, después, se enfrentaban los grupos.

(49) a. A la ruidosa, brutal y primitiva arremetida siguió pronto la táctica,

b. DESPUÉS CLS: 4 4 “los dos grupos corriendo para luchar” CAMBIAR OTRO MÉTODO.

c. Después, los dos grupos corriendo para luchar cambió a otro método.

(50) a. después se descubrió la emboscada y la falsa retirada

b. TERCERO, ESCONDERSE CLS: 4 1 “un grupo ataca a otro”. TAMBIÉN MÉTODO OTRO CLS: 4 1 “un grupo ataca a otro” CLS: 4 1 “un grupo se acerca a otro, se retira y, otra vez, vuelve a atacar”.

c. En tercer lugar, se escondían, un grupo se abalanzaba sobre otro, se retiraba y, después, volvía a atacar.

(51) a. y, luego, vino el orden de batalla con sofisticada distribución de armas ligeras y pesadas y empleo de caballería, carros de combate y elefantes.

b. TAMBIÉN ORDEN, CLP: 2 2 "personas (u objetos) en filas ordenadas de delante hacia atrás" ESPECIAL ARMA COSAS ETCÉTERA, EJEMPLO, PRIMERO LIGERO, AHÍ PESADO, CLP: 2 2 "personas (u objetos) en filas". USAR CABALLO+, CARRO, TAMBIÉN ELEFANTE.

c. También se ordenaban por filas las armas especiales, etcétera, por ejemplo, primero las ligeras, después las pesadas, las personas o los objetos se situaban en filas. Usaban caballos, carros y elefantes.

(52) a. En ningún campo fue tan fecundo el ingenio humano como en el de la invención de métodos estratégicos, de ataque y defensivos cada vez más eficaces.

b. TEMA INTELIGENCIA PERSONA+ HUMANO++ PARA CREAR MÉTODO TEMA ENFRENTARSE O DEFENDER OTRO TEMA IGUAL INTELIGENCIA VALE NO-HAY.

c. La inteligencia de las personas humanas para crear métodos de ataque o defensa no hay otro tema donde haya una inteligencia igual de válida.

(53) a. El objetivo ideal de la estrategia es la destrucción del enemigo.

b. OBJETIVO PERFECTO MÉTODO QUÉ-?, ENEMIGO DESAPARECER, OBJETIVO ESE.

c. El objetivo perfecto del método era hacer desaparecer al enemigo.

(54) a. La brutalidad puesta a ese servicio sólo conoce suavizamientos graduales según la época y el lugar.

b. MISMO BRUTALIDAD++ ESE, DEPENDER SUAVE O BRUTALIDAD DEPENDER ÉPOCA O ZONA GUERRA, DEPENDER.

c. La brutalidad o suavidad de eso dependía de la época o de la zona de la guerra.

(55) a. Cuando acaso se hacían prisioneros, su destino posterior era la esclavitud.

b. TOCA PERSONA PRESO, DESPUÉS, TRASLADAR CÁRCEL.

c. Cuando se hacían prisioneros se trasladaban a las cárceles.

(56) a. Las poblaciones sometidas conocieron siempre la ley de la espada.

b. GRUPO PERDER SOMETERSE CONOCER QUÉ-?, LEY ESPADA CONOCER, SOMETERSE.

c. Los grupos que perdían se sometían y conocían la ley de la espada.

(57) a. La historia de las guerras está llena de crueldad, y sangre en todas las partes del mundo.

b. HISTORIA TEMA GUERRA DENTRO SANGRE, BRUTALIDAD, MUNDO CLP: "se señalan las distintas partes en el mundo" TODO.

c. En la historia del tema de las guerras hay sangre y brutalidad en todas partes del mundo.

(58) a. Y poco ha cambiado, sigue habiendo guerras, hay matanzas de pueblos y nos llegan noticias de gran derramamiento de sangre.

b. AHORA-MISMO, DESDE-ANTES-HASTA-AHORA CAMBIAR POCO, CONTINUAR GUERRA, GRUPO++ ENFRENTARSE, RECIBIR-INFORMACIÓN NOTICIAS, RECIBIR-INFORMACIÓN, SANGRE, MATANZA, HORRIBLE, CONTINUAR++.

c. Desde antiguo hasta la actualidad ha cambiado poco, continúan las guerras y el enfrentamiento de los grupos, se reciben noticias de que continúan la sangre y las matanzas horribles.

(59) a. En lo que alcanza la historia mundial, hubo pueblos sometidos por otros pueblos belicosos y dispuestos a seguir en la agresión a caudillos ambiciosos, dominantes y brutales.

b. HISTORIA MUNDO HAY GRUPO SOMETERSE MOTIVO HAY ALGUNOS PODEROSO APLASTAR GRUPO SOMETERSE. ESE MISMO OBJETIVO QUÉ-?, PERSONA FIEL. ESE PERSONA CÓMO-?, GANAS GUERRA, SENTIR BRUTALIDAD ESE FIEL.

c. En la historia del mundo hay grupos que se someten porque hay algunos grupos poderosos que aplastan al grupo sometido, que tiene como objetivo ser fiel a esa persona. Esa persona es belicosa y brutal y le son fieles.

(60) a. El objetivo final era la conquista del mundo mediante la creación de imperios,

b. OBJETIVO FINAL QUÉ-?, MUNDO DOMINAR++, CREAR QUÉ-?, DL-IMPERIO.

c. El objetivo final es dominar el mundo y crear imperios.

(61) a. es lo que se conoce como imperialismo.

b. ESE NOMBRE CÓMO-?, DL-IMPERIALISMO.

c. Eso se llama imperialismo.

(62) a. Éste se define como un afán de dominio que procura extender la jurisdicción de un país a costa de países vecinos e incluso lejanos.

b. ESE QUÉ-? CONCEPTO SIGNIFICAR QUÉ-?, GANAS DOMINAR++, OBJETIVO INTENTAR MI PAÍS MI INTENTAR DOMINAR EXTENDERSE PAÍS++ EXTENDERSE, TAMBIÉN LEJOS PAÍS+ TAMBIÉN LEJOS.

c. Ese concepto significa ganas de dominar con el objetivo de intentar extender mi país también a países lejanos.

(63) a. La historia nos demuestra que la tendencia al imperialismo existe en todo tiempo y que en cuanto surge un tirano puede cristalizar enseguida.

b. HISTORIA DEMOSTRAR+ NORMAL ANTES INTENTAR OBJETIVO DOMINAR AHORA-NO, CONTINUAR, DESDE-ANTES-HASTA-AHORA CONTINUAR+. TOCA UN SURGIR PERSONA ORGULLOSO ÉL PODER CONTINUAR++.

c. La historia demuestra que el intento de dominar no es algo de ahora sino que continúa desde la antigüedad hasta ahora. Cuando surge una persona orgullosa puede continuar.

(64) a. Te voy a contar un día en la vida de mi perro.

b. YO CONTAR+ VIDA MI PERRO, UN DÍA COMPLETO YO CONTAR.

c. Yo voy a contar la vida de mi perro, un día completo.

(65) a. Por la mañana, mi perro se despierta a las 8.30.

b. POR-LA-MAÑANA LEVANTAR, DESPERTARSE, HORA OCHO Y-MEDIA DESPERTARSE.

c. Por la mañana se levanta, se despierta a las ocho y media.

(66) a. Él nunca se levanta sólo,

b. ÉL SÓLO LEVANTARSE NUNCA

c. Él nunca se levanta solo

(67) a. porque es muy dormilón.

b. MOTIVO DORMIR ENCANTAR.

c. porque le encanta dormir.

(68) a. Yo lo despierto para sacarlo a la calle.

b. YO LLAMAR PARA DESPERTAR CALLE CLI: “llevar el perro de la cadena”.

c. Yo lo llamo para despertarlo y sacarlo a la calle.

(69) a. Si hace mucho frío, le pongo su abrigo.

b. SI FRÍO, ABRIGO (cond)CLS: 2 “perro andando”.

c. Si hace frío le pongo un abrigo.

(70) a. En la calle se encuentra con otros perros.
b. TOCA CALLE NORMAL ENCONTRARSE+ PERRO,
c. En la calle normalmente se encuentra con perros,

(71) a. Muchos perros son amigos suyos.
b. ALGUNOS AMIGO SU.
c. algunos son amigos suyos.

(72) a. Cuando los ve, tira de la cadena para que lo acerque a sus amigos.
b. TOCA VER CLC: "perro que tira de la cadena" PARA CERCA CLS: 2 "perro andando".
c. Cuando los ve tira de la cadena para acercarse.

(73) a. Y, luego, se pone a mover el rabo y a jugar con ellos.
b. DESPUÉS, NORMAL RABO CLPC: 1 "mover el rabo" PARA JUGAR++.
c. Después, normalmente mueve el rabo para jugar.

(74) a. Después, volvemos a la casa y desayunamos.
b. DESPUÉS, TOCA IR CLS: 2 "los dos se marchan" CASA, TOCA DESAYUNO.
c. Después, los dos vamos a casa a desayunar.

(75) a. Mi perro no desayuna pero, a veces, le damos un trozo de tostada.
b. MI PERRO DESAYUNO NADA. MUCHAS-VECES TOSTADA ALGO CLI: "darle tostada al perro".
c. Mi perro no desayuna nada. Muchas veces le doy un trozo de tostada.

(76) a. Luego, todos salimos de casa:
b. DESPUÉS, TODOS IR CASA,
c. Después, todos salimos de casa,

(77) a. papá y mamá se van al trabajo
b. PADRE-MADRE IR TRABAJAR
c. mi padre y mi madre van a trabajar

(78) a. mi hermano y yo nos vamos al colegio.
b. HERMANO-HOMBRE YO IR COLEGIO.
c. mi hermano y yo vamos al colegio.

(79) a. Entonces, mi perro se queda solo.
b. DESPUÉS, PERRO SOLO CASA DENTRO.

c. Después, el perro se queda solo en casa.

(80) a. Cuando mi perro se queda solo, no se aburre.

b. TOCA ÉL SOLO ABURRIR (2m)NO.

c. Cuando está solo no se aburre.

(81) a. Duerme un poco pero también juega mucho.

b. ÉL DORMIR MUCHAS-VECES, JUGAR TAMBIÉN.

c. Duerme muchas veces, también juega.

(82) a. Mi perro tiene una caja con juguetes.

b. ÉL HAY CAJA DENTRO JUGUETE ETCÉTERA,

c. Tiene una caja con juguetes variados,

(83) a. Tiene tres pelotas de colores y varios muñecos de goma.

b. HAY PELOTA COLOR TRES, MÁS MUÑECO ESPECIAL GOMA.

c. tiene tres pelotas de colores y muñecos de goma.

(84) a. Cuando está solo, mi perro abre la caja con su hocico, saca todos sus juguetes y se pone a jugar.

b. TOCA ÉL SOLO, CAJA HOCICO CLPC: "levanta la tapa de la caja con su hocico" PARA JUGUETE (2m)FUERA++ JUGAR++.

c. Cuando se queda solo, abre la caja con su hocico para sacar los juguetes.

(85) a. Se pone a darle patadas a las pelotas y a correr tras ellas.

b. ÉL MUCHAS-VECES PELOTA CLPC: "dar patadas a las pelotas" CLS: 2 "el perro corre detrás de las pelotas".

c. Muchas veces da patadas a las pelotas y corre tras ellas.

(86) a. También le gusta morder los muñecos de goma.

b. TAMBIÉN MUCHAS-VECES MUÑECO ESPECIAL GOMA MORDER,

c. También muchas veces muerde los muñecos de goma.

(87) a. Se tira al suelo y se pasa las horas mordiéndolos.

b. SUELO CLL: 2 "tumbado en el suelo" HORA++ MORDER.

c. se tumba en el suelo y se pasa horas mordiendo.

(88) a. Él muerde sus muñecos con cuidado, nunca los rompe.

b. ÉL CLPC: "morder con fuerza" ROMPER MUÑECO-NO, ÉL CUIDAR.

c. Él no muerde con fuerza los muñecos, no los rompe, él los cuida.

(89) a. Así, pasa el tiempo hasta que volvemos a casa.
b. HORA++ HASTA TOCA TODOS IR CASA.
c. Se pasa las horas hasta que todos volvemos a casa.

(90) a. A las 14.30 volvemos a casa.
b. HORA DOS Y-MEDIA IR CASA.
c. A las dos y media vamos a casa.

(91) a. Antes de abrir la puerta mi perro ya sabe que hemos vuelto porque nos oye cuando llegamos al portal.
b. ANTES PUERTA ABRIR, ANTES SABER MOTIVO TOCA OÍR PORTAL ABAJO OÍR.
c. Antes de abrir la puerta lo sabe porque nos oye en el portal de abajo.

(92) a. Cuando abrimos la puerta nos encontramos a mi perro muy contento de vernos.
b. TOCA ABRIR, VER FELIZ, VER FELIZ.
c. Cuando abrimos lo vemos muy feliz.

(93) a. Enseguida, nos lavamos las manos y ponemos la mesa para almorzar.
b. DESPUÉS, DIRECTO IR LAVARSE-LAS-MANOS, MÁS MESA PREPARAR PLATO++ ETCÉTERA.
c. Después, vamos directo a lavarnos las manos y a poner la mesa.

(94) a. Durante el almuerzo mi perro se tumba debajo de la mesa.
b. MESA COMIDA TOCA CLL: 2 “el perro se pone debajo de la mesa”.
c. En la comida, el perro se pone debajo de la mesa.

(95) a. Mamá no quiere que le demos de comer.
b. MAMÁ QUERER-NO COMIDA DAR QUERER-NO.
c. Mamá no quiere que le demos comida.

(96) a. Mi perro espera debajo de la mesa por si se cae algo.
b. ÉL ESPERAR A-VER SUERTE HAY DAR-ME COMIDA ALGO+.
c. Él espera a ver si hay suerte y le damos algo de comida.

(97) a. Tras el almuerzo, le damos su comida.
b. FIN COMIDA, DESPUÉS, TURNO COMER.
c. Tras la comida, después, le toca comer

(98) a. Mi perro come pienso y croquetas para perro.

b. SU ESPECIAL COMIDA QUÉ-?, ESPECIAL COMIDA ANIMALES PERRO+, MÁS CROQUETA ESPECIAL PERRO.

c. su comida especial para perros y croquetas para perros.

(99) a. Después, saco a mi perro otra vez a la calle.

b. DESPUÉS IR CALLE OTRA-VEZ.

c. Después, vamos a la calle otra vez.

(100) a. Esta vez se lleva en la boca su pelota favorita.

b. ÉL IR QUÉ-HACER-?, PELOTA GUSTAR SU COGER, IR.

c. Él coge la pelota que le gusta.

(101) a. En la calle jugamos durante un buen rato.

b. LOS-DOS JUGAR+,

c. Los dos jugamos,

(102) a. Primero suelto a mi perro y después le tiro la pelota muy lejos.

b. PRIMERO YO CLI: "desatar al perro", PELOTA TIRAR LEJOS,

c. primero yo desato al perro, le tiro la pelota lejos,

(103) a. Él corre muy deprisa tras la pelota, la coge y me la trae.

b. CORRER CLS: 2 "el perro corre detrás de las pelotas" COGER CLS: 2 "el perro regresa corriendo".

c. corre tras la pelota, la coge y regresa corriendo.

(104) a. Y, otra vez, se la vuelvo a tirar.

b. PERRO OTRA-VEZ DAR-ME, TIRAR LEJOS.

c. El perro otra vez me la da, la tiro lejos.

(105) a. A mi perro le gusta mucho este juego.

b. ÉL JUGAR ENCANTAR,

c. Le encanta jugar,

(106) a. Puede pasarse horas corriendo detrás de la pelota.

b. CAPAZ HORA++ PELOTA CLS: 2 "el perro corre detrás de las pelotas" CAPAZ.

c. es capaz de pasarse horas corriendo detrás de la pelota.

(107) a. A las 17.00 volvemos a casa.

b. HORA CINCO OTRA-VEZ CLS: 2 “los dos se marchan” CASA.

c. A las cinco otra vez nos marchamos a casa.

(108) a. Yo me pongo a hacer los deberes.

b. YO COLEGIO PREPARAR DEBERES.

c. Yo preparo los deberes del colegio.

(109) a. Mi perro se vuelve a dormir porque está muy cansado de correr.

b. ÉL OTRA-VEZ DORMIR MOTIVO CANSADO ANTES CORRER,

c. Otra vez duerme porque está cansado de correr,

(110) a. Él duerme hasta la hora de merendar.

b. AHORA DORMIR HASTA HORA COMER MEDIA TARDE ÉL.

c. ahora duerme hasta la hora de merendar.

(111) a. Después de merendar, yo me voy a la calle a jugar con mi hermano y nuestros amigos.

b. DESPUÉS COMER TARDE IR JUGAR AMIGO+ ETCÉTERA MÁS HERMANO-HOMBRE, YO, LOS-DOS IR.

c. Después de merendar voy a jugar con los amigos y mi hermano.

(112) a. Mi perro se queda ahora en casa jugando con sus pelotas.

b. PERRO QUÉ-HACER-?, CASA JUGAR PELOTA JUGAR++.

c. El perro se queda en casa jugando con la pelota.

(113) a. Tras la cena, le toca a papá sacarlo a la calle.

b. COMER NOCHE FIN, PAPÁ RESPONSABLE CALLE CLI: “llevar al perro de la cadena”.

c. Tras la cena, papá es responsable de sacar al perro a la calle.

(114) a. Ya está muy oscuro

b. CIELO OSCURO

c. Está oscuro

(115) a. y mamá no quiere que los niños salgamos de casa.

b. MOTIVO MAMÁ QUERER-NO NIÑOS IR CALLE QUERER-NO MOTIVO OSCURO.

c. y mamá no quiere que los niños vayamos a la calle porque está oscuro.

(116) a. Luego, cuando sube de la calle, mi perro se acuesta en su cama.

b. DESPUÉS TOCA CASA, ÉL CAMA DORMIR.

c. Después, en casa, se duerme en la cama.

(117) a. De madrugada, se levanta a beber agua, pero la mayor parte de la noche se la pasa durmiendo.

b. MUCHAS-VECES MADRUGADA IR AGUA BEBER, NORMAL DÍA POR-LA-NOCHE COMPLETO DORMIR.

c. Muchas veces, de madrugada, va a beber agua, normalmente duerme toda la noche.

(118) a. Actualmente la población mundial crece a un ritmo lo suficientemente acelerado, rápido y vertiginoso como para temer consecuencias catastróficas.

b. AHORA-MISMO PERSONA+ MUNDO CRECER VELOZ, RÁPIDO, BURRADA, PODER FUTURO PROBLEMA+.

c. En la actualidad la población del mundo crece velozmente, rápidamente, una burrada, puede haber problemas en el futuro.

(119) a. Las Naciones Unidas predicen que en el año 2050 el planeta acogerá a 8.900 millones de personas, aunque la cifra podría llegar a los 10.700 millones.

b. AHORA-MISMO DL-ONU DECIR 2050 MUNDO TOTAL PERSONA+ 8900000000 PERSONA+, PERO CAPAZ MÁXIMO 10700000000 PERSONA+ VIVIR MUNDO.

c. En la actualidad la ONU dice que en el 2.050, en el mundo habrá un total de 8.900 millones de personas, pero que se puede llegar a un máximo de 10.700 millones de habitantes del mundo.

(120) a. A pesar de todo, hay que decir que, en general, las mujeres están teniendo menos hijos que nunca

b. PODER DECIR EN-GENERAL MUJER+ HIJO MUY-POCO, ANTES MÁS.

c. En general, se puede decir que las mujeres están teniendo muy pocos hijos, antes tenían más.

(121) a. y que el crecimiento de la población está siendo más lento, pausado y menor de lo que se esperaba.

b. PERSONA CRECER LENTO, TRANQUILO, DESACELERADO.

c. La población crece lentamente, tranquilamente, de manera desacelerada.

(122) a. Esto puede deberse a que las políticas de control de nacimientos han reducido la natalidad más de lo que se esperaba

b. ESE MOTIVO POLÍTICA CONTROL NACIMIENTO++ INTENTAR EVITAR, NACIMIENTO++-NO, EVITAR MÁS INTENTAR.

c. Eso es por la política de control de nacimientos que intenta evitar los nacimientos.

(123) a. y a que las muertes debido al SIDA también han sido mayores de lo esperado.

b. ADEMÁS PERSONA+ MUERTO+ MOTIVO SIDA MÁS CRECER.

c. Además las personas muertas por el SIDA han crecido más.

(124) a. Hasta el año 2020, el crecimiento de la población mundial será de 64 millones de personas al año

b. HASTA TOCA AÑO 2020 CRECER EN-TOTAL EN-UN-AÑO 64000000 PERSONA+ EN-UN-AÑO.

c. Hasta el año 2020 crecerá en un año, en total, 64 millones de personas al año.

(125) a. (en la actualidad la cifra está en torno a los 78 millones).

b. EN-ESTE-MOMENTO, AHORA-MISMO, NÚMERO CIFRA 78000000 PERSONA+ EN-UN-AÑO.

c. En este momento, en la actualidad, la cifra es de 78 millones de personas en un año.

(126) a. En el año 2050 se estima que la población crecerá en 33 millones de personas por año.

b. TOCA AÑO 2050, MÁS-O-MENOS+, ESTIMAR, MÁS-O-MENOS, 33000000 PERSONA+ EN-UN-AÑO.

c. En el año 2050, más o menos, se estima 33 millones de personas en un año.

(127) a. En definitiva, se prevé que la población no crecerá tanto como estaba previsto en un principio.

b. EN-RESUMEN, MÁS-O-MENOS+, PERSONA+ CLD: C C "crecer cada vez más y muy rápidamente"-NO. AL-PRINCIPIO PENSAR CLD: C C "crecer cada vez más y muy rápidamente", AHORA VER LENTO MENOS++.

c. En resumen, más o menos, la población crece exageradamente. Al principio se pensaba en un crecimiento exagerado, ahora se ve que es lento, que es menor.

(128) a. Las tasas más altas de crecimiento de la población corresponden a los países más pobres, deprimidos y subdesarrollados.

b. MÁS-O-MENOS EN-TOTAL NÚMERO CIFRA PERSONA+ NORMALMENTE PAÍS++ QUIÉN-?, PAÍS+ POBRE, DESARROLLO NO-HAY, PAÍS+ DEPRIMIDO.

c. Más o menos, en total, la cifra de personas normalmente se da en los países pobres, subdesarrollados, deprimidos.

(129) a. Por contraste, en 61 países, donde reside casi la mitad de la población mundial, las parejas están teniendo menos de 2 hijos por mujer, cantidad necesaria para que haya un reemplazamiento generacional.

b. EN-CAMBIO, COMPARACIÓN+, 61 PAÍS++ CASI DENTRO VIVIR PERSONA MITAD PERSONA+ VIVIR MUNDO,PAREJA+ NORMALMENTE, HIJO DOS-NO, MENOS. UN MUJER SABER DOS HIJOS DEBER PARA EVOLUCIÓN CONTINUAR+ DEBER.

c. En cambio, en comparación, en 61 países donde viven casi la mitad de las personas del mundo, las parejas normalmente tienen menos de dos hijos. Se sabe que una mujer debe tener dos hijos para que la evolución continúe.

(130) a. En América del Norte, Japón y Europa, el crecimiento de la población prácticamente se ha detenido.

b. EJEMPLO AMÉRICA NORTE, JAPÓN, TAMBIÉN EUROPA PERSONA+ CLD: C C "crecer cada vez más y muy rápidamente"-NO, DETENER, PARAR.

c. Por ejemplo, en América del Norte, Japón y Europa el crecimiento de la población no es exagerado, se ha parado, se ha detenido.

(131) a. Cada mujer de los países de la Unión Europea tiene 1,4 hijos.

b. UN MUJER PAÍS+ UNIÓN EUROPEA EN-TOTAL HIJO 1,4 HIJO UN MUJER.

c. En los países de la Unión Europea, en total, cada mujer tiene 1,4 hijos.

(132) a. España está a la cabeza de los países con la tasa de fecundidad más baja, con un promedio de 1,15 hijos por mujer.

b. ESPAÑA, PAÍS ESPAÑA, PRIMERO NACIMIENTO++ MUY-POCO, MÁS-O-MENOS, 1,15 HIJO UN MUJER.

c. España es el primer país en baja tasa de natalidad, más o menos, 1,15 hijos por mujer.

(133) a. En España las tasas de natalidad están descendiendo desde hace 20 años, circunstancia que no se ha dado en ningún otro país de Europa.

b. ESPAÑA EN-TOTAL NACIMIENTO++ EN-UN-AÑO DESDE-TIEMPO-HASTA-AHORA 20 AÑO DESCENDER, IGUAL PAÍS OTRO PAÍS EUROPA NO-HAY, ÚNICO ESPAÑA.

c. En España la tasa anual de natalidad desde hace 20 años hasta ahora ha descendido, no hay otro país así en toda Europa, el único es España.

(134) a. Italia y Japón son los países que tienen la población más anciana, pero en el 2050 está previsto que este puesto lo ocupe España.

b. ITALIA, JAPÓN, ESOS PAÍS+ AHORA-MISMO EDAD TERCERO MÁS, PERO TOCA 2050 QUIÉN-? CABEZA, PRIMERO, QUIÉN-?, ESPAÑA.

c. En la actualidad Italia y Japón tienen más personas de la tercera edad, pero cuando llegue el 2.050 España será el primero, estará a la cabeza.

(135) a. Las razones por las que las españolas tienen tan pocos hijos parecen ser:

b. MOTIVO PORQUE MUJER+ HIJO POCO MOTIVO-?:

c. Los motivos porque las mujeres tienen tan pocos hijos son:

(136) a. El descenso en el número de casamientos.

b. AHORA-MISMO CASAMIENTO++ REDUCIR.

c. En la actualidad se han reducido los casamientos.

(137) a. A pesar de que ha aumentado el número de personas que conviven juntas sin estar casadas, no nacen apenas niños de estas parejas.

b. APARTE HAY PAREJA-DE-HECHO CRECER, NO-HACER-FALTA CASAR NO-HACER-FALTA, PERO PAREJA-DE-HECHO NO-HAY HIJO, NO-HAY.

c. Por otra parte, las parejas de hecho han crecido, no hace falta casarse, pero las parejas de hecho no tienen hijos.

(138) a. Las cifras de paro.

b. ADEMÁS PROBLEMA ESTADÍSTICA PARO.

c. Además el problema de las estadísticas de paro.

(139) a. En España se da una de las tasas de desempleo entre los jóvenes más altas de Europa y aquellos que consiguen un trabajo éste suele ser precario, inestable y mal pagado.

b. ESPAÑA PROBLEMA PARO JOVEN, PERO EUROPA PRIMERO PARO JOVEN QUIÉN-? EUROPA, ADEMÁS PERSONA+ JOVEN ENCONTRAR TRABAJO PROBLEMA MOTIVO TRABAJO PRECARIO, MÁS TEMPORAL TRABAJO, DINERO GRAN-SUELDO-NO, SUELDO-BAJO.

c. En España el problema del paro juvenil es el más alto de Europa, además, los jóvenes que encuentran trabajo tienen el problema de que éste es precario, temporal y con bajo sueldo.

(140) a. El desempleo afecta más a las mujeres.

b. PROBLEMA TRABAJO MUJER PARO MÁS MUJER.

c. El problema del trabajo femenino. Hay más paro en la mujer.

(141) a. El empresario tiene miedo a que las mujeres se queden embarazadas

b. PERSONA EMPRESA PROPIETARIO MIEDO MOTIVO MUJER EMBARAZO,

c. Los empresarios tienen miedo porque la mujer se quede embarazada,

(142) a. y tengan que pedir permiso de maternidad.

b. ADEMÁS SABER DESPUÉS DEBER PEDIR PARA EMBARAZO CUIDAR POR-ESO PROBLEMA.

c. además saben que después deben pedir la baja maternal por eso es un problema.

(143) a. Por ello, muchas mujeres renuncian a tener hijos.

b. MOTIVO MUCHAS-VECES MUJER DECIR HIJO EMBARAZO NACIMIENTO NADA, RECHAZAR.

c. Por ello, muchas veces las mujeres dicen no tener hijos, lo rechazan.

(144) a. El coste de los hijos es muy alto. Para mantener un nivel de consumo determinado es necesario que trabajen los dos miembros de la pareja,

b. ADEMÁS HIJO CUIDAR MUY-CARO, PARA MANTENER BIEN, DEBER DOS PAREJA, DOS, HOMBRE, MUJER, TRABAJAR DOS.

c. Además, el cuidado de los hijos es muy caro, para mantenerlos bien los dos de la pareja, el hombre y la mujer, deben trabajar los dos.

(145) a. lo que muchas veces hace difícil que haya una conciliación entre trabajo y familia.

b. ENTONCES, MUCHAS-VECES PROBLEMA QUÉ-?, TRABAJO, FAMILIA UNIR BIEN PROBLEMA.

c. Entonces, muchas veces existe el problema de unir bien familia y trabajo.

(146) a. Mientras que en España el peso de los hijos repercuta más sobre la mujer no se podrá esperar un aumento de la fecundidad.

b. ENTONCES, ESPAÑA EN-ESTE-MOMENTO HIJO RESPONSABLE MÁS QUIÉN-? MUJER+, POSIBLE FUTURO MUJER NACIMIENTO++ CRECER-NO.

c. En España, en la actualidad, son las mujeres las máximas responsables de los hijos, puede que en el futuro la natalidad no crezca.

(147) a. El año pasado, unas treinta mujeres murieron en España a manos de sus maridos.

b. AÑO PASADO EN-TOTAL 30 MUJER+ MUERTO MOTIVO HOMBRE MARIDO SU MATAR.

c. El año pasado, en total, murieron 30 mujeres asesinadas por sus maridos.

(148) a. Además, se tramitaron más de 16.000 denuncias por malos tratos de hombres a sus parejas.

b. ADEMÁS, EN-TOTAL 16000 DENUNCIA MUJER DENUNCIA MARIDO PAREJA SU.

c. Además, en total, hubo 16.000 denuncias de mujeres a sus maridos o parejas.

(149) a. Sin embargo, los especialistas dicen que se dan muchas más agresiones a las mujeres de las que se denuncian.

b. PERO, PERSONA+ AUTORIDAD++ DECIR NORMAL MATAR, MALTRATO MÁS, DENUNCIA MENOS.

c. Pero, las autoridades dicen que, normalmente hay más asesinatos, maltratos que denuncias.

(150) a. Los españoles han visto en diez días dos casos terribles.

b. PERSONA+ ESPAÑA VIVIR EN-TOTAL DESDE-HACE-TIEMPO-HASTA-AHORA DIEZ DÍA DOS EJEMPLO HORRIBLE.

c. Los españoles en diez días (han visto) dos ejemplos horribles.

(151) a. En uno de ellos, una mujer moría a manos de su ex-novio, un preso peligroso en libertad.

b. UN MUJER MUERTO MOTIVO EX NOVIO ANTES HOMBRE DENTRO CÁRCEL, AHORA LIBRE.

c. Una mujer asesinada por su ex-novio que antes estaba en la cárcel y ahora estaba libre.

(152) a. La mujer había realizado ya muchas denuncias por malos tratos.

b. ESE MUJER MUCHAS-VECES ANTES DENUNCIA+ MOTIVO MALTRATO.

c. Esa mujer había denunciado muchas veces antes el maltrato.

(153) a. En el otro caso, el marido arrojó a su mujer por el balcón de un séptimo piso.

b. OTRO SEGUNDO MARIDO HOMBRE SABER BALCÓN SÉPTIMO-PISO TIRAR.

c. En el segundo, el marido la tiró desde el balcón de un séptimo piso.

(154) a. Ante esto, muchas personas nos preguntamos por qué suceden estos crueles actos de violencia.

b. ENTONCES, MUCHAS-VECES PERSONA PENSAR, VER PORQUÉ TEMA MALTRATO PORQUÉ.

c. Entonces, muchas veces las personas piensan porqué se da el tema del maltrato, porqué.

(155) a. A pesar de los datos anteriores, no podemos decir que el número de malos tratos haya aumentado en nuestra época;

b. ANTES DATO++ DECIR EN-TOTAL NÚMERO MALTRATO AHORA MÁS-NO, NO, NO-PODER CONTAR, NO-PODER,

c. Los datos sobre el número total de malos tratos no son más ahora que antes, no se puede contar eso,

(156) a. lo que ocurre es que la sociedad es más sensible y está más informada.

b. QUÉ-PASA-?, AHORA-MISMO SOCIEDAD SENSIBLE, ADEMÁS RECIBIR-INFORMACIÓN++ CLARO POR-ESO.

c. ¿qué es lo que pasa?, que en la actualidad la sociedad es más sensible, además se recibe una información clara, por eso.

(157) a. Este problema se ha dado en todas las épocas, culturas y clases sociales.

b. ESE PROBLEMA MALTRATO TODOS ÉPOCA DURANTE HISTORIA DURANTE, CULTURA, ETCÉTERA, ADEMÁS GRUPO+++.

c. El problema del maltrato se ha dado en todas las épocas de la historia, en todas las culturas y grupos.

(158) a. No es que ahora hayan aumentado, lo que ocurre es que se denuncia más que antes.

b. AHORA ESPECIAL MÁS-NO, AHORA QUÉ-PASA-?, DENUNCIA.

c. Ahora no se da especialmente más, lo que pasa ahora es que se denuncia.

(159) a. Ahora, la mujer que denuncia se siente más apoyada.

b. AHORA MUJER TOCA DENUNCIA HAY APOYO,

c. Ahora se apoya a la mujer que denuncia,

(160) a. Antes, las mujeres que se culpaban a sí mismas de recibir malos tratos, se consideraban malas esposas.

b. ANTES MUJER SENTIR CULPA MÍ, SENTIR MAL, NORMAL MUJER PARECER MUJER ESPOSA MALA, POR-ESO.

c. antes la mujer se sentía culpable, se sentía mal, normalmente la mujer se consideraba mala esposa por eso.

(161) a. Los factores que influyen en la violencia en el hogar son múltiples.

b. ENTONCES, HAY DATO+++ PARA CASA MALTRATO INFLUENCIA QUÉ-? ETCÉTERA.

c. Hay datos que influyen en el maltrato en el hogar.

(162) a. Se sabe que aquellas personas que durante la infancia han recibido malos tratos es más probable que sean agresivas y violentas en sus familias.

b. SABER TOCA PERSONA HIJO PEQUEÑO DECIR RECIBIR-MALTRATO NORMAL CRECER, DESPUÉS, CRECER, FAMILIA NORMAL RECIBIR-MALTRATO CAPAZ FAMILIA MALTRATAR++, CAPAZ.

c. Se sabe que cuando una persona es pequeña y recibe maltrato, normalmente crece y, después de crecer en una familia recibiendo maltrato es capaz de maltratar a su familia.

(163) a. Los malos tratos se dan en todas las clases sociales,

b. NORMAL MALTRATO TODOS GRUPO++ SOCIEDAD TODOS IGUAL HAY.

c. Normalmente el maltrato se da en todos los grupos sociales por igual.

(164) a. aunque se ha podido comprobar que a menores recursos económicos hay más violencia.

b. DECIR DINERO POBRE, POCO, NORMAL MALTRATO MÁS.

c. Si se es pobre, si se tiene poco dinero, normalmente se da más maltrato.

(165) a. Cuando la situación económica es buena es más fácil separarse, pero si la mujer depende económicamente del hombre y teme no tener dinero para alimentar a sus hijos es probable que aguante las agresiones.

b. EJEMPLO SITUACIÓN DINERO BIEN PODER FÁCIL SEPARARSE, PERO MUJER EJEMPLO HOMBRE MARIDO DEBER DINERO MANTENER++ DESPUÉS, PROBLEMA SEPARARSE, HIJO, NORMAL MUJER QUÉ-HACER-?, AGUANTAR MALTRATO, AGUANTAR.

c. Si la situación económica es buena, puede separarse fácilmente, pero si el marido debe mantener a la mujer, después está el problema de separarse, de los hijos. Normalmente, la mujer ¿qué hace?, aguanta el maltrato.

(166) a. Por lo general, la violencia en la pareja es más frecuente entre personas jóvenes, aunque las denuncias lleguen después de varios años de matrimonio.

b. EN-GENERAL DENTRO PAREJA VIOLENCIA USAR MÁS JOVEN, PODER DENUNCIA DESPUÉS MATRIMONIO, DESPUÉS DENUNCIA.

c. En general, dentro se suele dar más maltrato en las parejas de jóvenes, pero puede que la denuncia se dé después del matrimonio.

(167) a. Los expertos afirman que los primeros malos tratos ya se manifiestan desde el noviazgo, lo que ocurre es que inmediatamente después de la agresión viene el arrepentimiento, acompañado de las promesas y de las muestras de amor, que en la mayoría de las veces hacen que la víctima perdone.

b. PERSONA AUTORIDAD++ DECIR MALTRATO EMPEZAR CUANDO NOVIO+ EMPEZAR MALTRATO, PERO QUÉ-PASA-?, DESPUÉS MALTRATO PERSONA ARREPENTIRSE ADEMÁS PEDIR-PERDÓN, ADEMÁS PEDIR-PERDÓN, DECLARACIÓN-DE-AMOR, MUCHAS-VECES QUÉ-PASA-?, MUJER PERDONAR

c. Las autoridades dicen que el maltrato empieza cuando son novios, pero ¿qué pasa?, después del maltrato la persona se arrepiente, pide perdón y declara su amor. Muchas veces lo que pasa es que la mujer perdona.

(168) a. También es importante señalar que factores como el consumo de alcohol o la contaminación pueden favorecer el maltrato.

b. TAMBIÉN IMPORTANTE DECIR COSAS EJEMPLO BEBIDA ALCOHOL O CONTAMINACIÓN DAR-ME FÁCIL TRASTORNAR, MALTRATO

c. También es importante decir que cosas como las bebidas alcohólicas o la contaminación pueden fácilmente trastornar y (provocar) el maltrato.

(169) a. El ritmo de vida de las grandes ciudades que favorece la excitabilidad y el estrés también pueden causar una mayor violencia en el hogar.

b. AHORA-MISMO CIUDAD GRANDE EVOLUCIÓN DAR-ME+ FÁCIL AGRESIVIDAD ADEMÁS ESTRÉS DAR-ME++ ADEMÁS CAPAZ DENTRO CASA MALTRATO MÁS.

c. En la actualidad la evolución de las ciudades grandes facilita la agresividad y puede provocar estrés y puede que en el hogar haya más maltrato.

(170) a. Aunque el hecho de que haya más denuncias en las ciudades que en los pueblos no quiere decir que en estos últimos haya un número menor de malos tratos.

b. PERO AHORA-MISMO CIUDAD DENUNCIA MÁS, PUEBLO MENOS, QUERER DECIR DENTRO PUEBLO, DENTRO, MALTRATO MENOS-?, NO++, DEPENDER.

c. Pero que en la actualidad se denuncie más en la ciudad que en los pueblos ¿quiere decir que en los pueblos hay menos maltrato? No, depende.

(171) a. Los datos pueden que no nos estén dando un auténtico reflejo de la realidad.

b. MUCHAS-VECES DATO++ DECIR DAR-ME INFORMACIÓN REAL-NO.

c. Muchas veces los datos no nos dan una información real.

(172) a. En las grandes ciudades la mujer puede hacer la denuncia sin que se sepa.

b. CIUDAD GRANDE MUJER IR DENUNCIA, SECRETO, DIVULGAR NO-HAY.

c. En las ciudades grandes la mujer va a denunciar y hay secreto, no se divulga.

(173) a. En los pueblos y las ciudades pequeñas le puede costar más ir a un centro de salud o a la policía,

b. PERO PUEBLO O CIUDAD PEQUEÑO COSTAR IR CENTRO SALUD O POLICIA, IR COSTAR

c. Pero en los pueblos o en las ciudades pequeñas cuesta ir al centro de salud o a la policía

(174) a. porque es probable que la conozcan a ella, a su marido, a la familia de ambos, etc.

b. MOTIVO POSIBLE CONOCER PERSONA MUJER O HOMBRE MARIDO O FAMILIA LOS-DOS, POSIBLE, ETCÉTERA.

c. porque es posible que conozcan a la mujer o al marido o a la familia de ambos, etcétera.

(175) a. Te voy a contar un día en la vida de mi hermana.

b. YO CONTAR VIDA MI HERMANO-MUJER. YO CONTAR UN DÍA COMPLETO.

c. Te voy a contar la vida de mi hermana. Te voy a contar un día completo.

(176) a. Mi hermana se llama Ana y tiene 5 años.

b. ÉL NOMBRE DL-ANA, EDAD CINCO.

c. Ella se llama Ana. Tiene cinco años.

(177) a. Por la mañana, Ana se despierta a las 8.30.

b. POR-LA-MAÑANA DESPERTARSE HORA OCHO Y-MEDIA.

c. Por la mañana se despierta a las ocho y media.

(178) a. Ella nunca se levanta sola, porque es muy dormilona.

b. ÉL LEVANTARSE SOLO NUNCA MOTIVO+ ENCANTAR DORMIR.

c. Ella nunca se levanta sola porque le encanta dormir.

(179) a. Cuando suena el despertador, ella nunca lo oye.

b. TOCA DESPERTADOR ESCUCHAR PASAR+,

c. No oye el despertador.

(180) a. Yo la despierto para que se lave antes de desayunar.

b. YO CLS: 2 "persona andando" LLAMAR PARA LAVARSE ANTES DESAYUNO.

c. Yo la llamo para que se lave antes de desayunar.

(181) a. Si hace mucho frío, le llevo la bata a la cama.
b. EJEMPLO FRÍO YO LLEVAR BATA CAMA LLEVAR,
c. Si hace frío le llevo la bata a la cama.

(182) a. Es una bata rosa con un oso dibujado en el bolsillo.
b. COLOR ROSA, BOLSILLO OSO DIBUJO CLL: "dibujo sobre el bolsillo".
c. Es de color rosa. Tiene un dibujo de un oso en el bolsillo.

(183) a. A Ana le gusta mucho su bata.
b. ÉL MUJER("Ana") ENCANTAR BATA.
c. A Ana le encanta la bata.

(184) a. Después de salir del baño, desayunamos.
b. FIN SERVICIO, FIN, LOS-DOS DESAYUNO.
c. Cuando termina en el servicio, las dos vamos a desayunar.

(185) a. A Ana le gusta desayunar despacio.
b. ÉL MUJER("Ana") ENCANTAR DESPACIO DESAYUNO DESPACIO,
c. A Ana le encanta desayunar despacio.

(186) a. A veces mamá le dice que se dé prisa porque va a llegar tarde a clase.
b. MUCHAS-VECES MAMÁ(rol:madre): POR-FAVOR, HORA MUY-TARDE, CLASE, VAMOS DEPRISA.
c. Muchas veces mamá le dice: por favor, es tarde para la clase, date prisa.

(187) a. Ella desayuna todos los días leche caliente y una tostada con aceite.
b. ÉL TODOS-LOS-DÍAS LECHE CALIENTE ADEMÁS TOSTADA ACEITE, TODOS-LOS-DÍAS DESAYUNO.
c. Ella todos los días desayuna leche caliente y tostada con aceite.

(188) a. Los domingos papá compra churros y,
b. DOMINGO+ PAPÁ CHURRO COMPRAR,
c. Los domingos papá compra churros

(189) a. entonces, Ana cambia su desayuno.
b. ÉL CAMBIAR DESAYUNO, CAMBIAR.
c. ella cambia el desayuno.

(190) a. Tras desayunar, Ana y yo cogemos nuestras mochilas y nos vamos al colegio.

b. FIN DESAYUNO LOS-DOS IR PONERSE-LA-MOCHILA IR COLEGIO.

c. Después de desayunar, las dos nos ponemos las mochilas y nos vamos al colegio.

(191) a. Papá nos lleva a la esquina de la calle.

b. PAPÁ ACOMPAÑAR ESQUINA CALLE,

c. Papá nos acompaña a la esquina de la calle.

(192) a. Allí nos recoge el autobús del colegio.

b. ESQUINA, AHÍ VENIR AUTOBÚS. AUTOBÚS ESPECIAL COLEGIO.

c. En la esquina viene el autobús del colegio.

(193) a. Cuando llega el autobús, todos los niños se ponen en fila. Los menores entran los primeros.

b. TOCA VENIR AUTOBÚS TODOS NIÑOS CLP: 5 5 "fila de niños". EJEMPLO PEQUEÑO PRIMERO CLP: 5 5 "fila de niños",

c. Cuando viene el autobús todos los niños nos ponemos en fila, los más pequeños los primeros.

(194) a. Ana siempre entra antes que yo.

b. MUJER("Ana") PRIMERO, YO SEGUNDO.

c. Primero va Ana y en segundo lugar yo.

(195) a. Cuando llegamos al colegio, Ana se va a su clase.

b. ENTONCES, TOCA COLEGIO MUJER("Ana") IR CLASE AULA SU.

c. En el colegio Ana se va a su clase.

(196) a. Yo no la vuelvo a ver hasta la hora del recreo.

b. DESPUÉS, VERSE HORA DESCANSO VERSE.

c. Después nos vemos en la hora del recreo.

(197) a. Ana ya sabe leer y escribir. Es muy lista.

b. ÉL MUJER("Ana") SABER LEER, ESCRIBIR, MUY-INTELIGENTE.

c. Ana sabe leer y escribir, es muy inteligente.

(198) a. La maestra dice que es una niña muy buena, aunque algo traviesa.

b. MUJER-PROFESOR DECIR MUY BUENO, PERO UN-POCO TRAVIESO++.

c. La profesora dice que es muy buena, pero un poco traviesa.

(199) a. En clase, Ana aprende muchas cosas. A ella le gusta mucho dibujar y leer.

b. DENTRO CLASE MUJER(“Ana”) APRENDER MUCHO COSAS ETCÉTERA, ENCANTAR DIBUJAR, LEER,

c. En la clase Ana aprende muchas cosas diferentes, le encanta dibujar, leer,

(200) a. También le gusta aprender cosas sobre la vida de los animales.

b. TAMBIÉN GUSTAR COSAS VIDA SU ANIMALES GUSTAR.

c. también le gustan las cosas de la vida de los animales.

(201) a. A las 14.30 volvemos a casa.

b. HORA DOS Y-MEDIA, IR CASA.

c. A las dos y media vamos a casa.

(202) a. Nos lavamos las manos

b. LAVARSE-LAS-MANOS.

c. Nos lavamos las manos

(203) a. ayudamos a mamá y a papá a poner la mesa.

b. TAMBIÉN PADRE-MADRE MESA PREPARAR PLATO.

c. también papá y mamá ponen los platos en la mesa.

(204) a. Tenemos mucha hambre

b. HAMBRE MUCHO,

c. Tenemos mucha hambre

(205) a. nos comemos todo lo que nos ponen.

b. COMER PLATO AGOTADO.

c. comemos todo el plato.

(206) a. Ana bebe zumo de naranja mientras almuerza.

b. BEBER ZUMO NARANJA COMER.

c. Bebe zumo de naranja con la comida.

(207) a. También le gusta comer pan blanco.

b. TAMBIÉN, GUSTAR QUÉ-?, PAN BLANCO,

c. También le gusta el pan blanco

(208) a. Su plato favorito es la tortilla de patatas y el arroz blanco.

b. ADEMÁS ENCANTAR ARROZ BLANCO, TORTILLA PATATA.

c. además le encanta el arroz blanco y la tortilla de patatas.

(209) a. Después de la comida, mamá siempre nos da un poco de chocolate.
b. DESPUÉS COMER FIN COMER, CHOCOLATE MAMÁ DAR-ME.
c. Después de comer, mamá nos da chocolate.

(210) a. Tras el almuerzo, Ana y yo quitamos la mesa y ayudamos a poner los platos en el lavavajillas.
b. FIN COMIDA MEDIODÍA MUJER("Ana") LOS-DOS MESA PLATO CLI: "quitar los platos de la mesa" TAMBIÉN APOYAR PARA LAVAVAJILLAS PLATO CLI: "poner los platos en el lavavajillas".
c. Tras el almuerzo las dos quitamos la mesa y ayudamos a poner los platos en el lavavajillas.

(211) a. Tras esto, nos sentamos con papá y mamá en el salón.
b. FIN, SALÓN SENTAR PADRE-MADRE.
c. Cuando terminamos nos sentamos en el salón con papá y mamá.

(212) a. Unos días papá pone el vídeo y vemos películas de animales o de dibujos animados.
b. ALGUNOS DÍA+ VER VÍDEO ANIMADO DIBUJO O ANIMALES.
c. Algunos días vemos vídeos de dibujos animados o animales.

(213) a. Otros días, jugamos a distintos juegos.
b. OTRO DÍA, DEPENDER, JUGAR COSAS ETCÉTERA, DEPENDER.
c. Otros días jugamos a distintas cosas, depende.

(214) a. A las 17.30 nos ponemos a hacer los deberes.
b. HORA CINCO Y-MEDIA LOS-DOS ESPECIAL COLEGIO DEBERES PREPARAR.
c. A las cinco y media preparamos los deberes del colegio.

(215) a. Ana casi nunca tiene que hacer nada.
b. MUJER("Ana") HACER NADA.
c. Ana no hace nada.

(216) a. Entonces, se pone a leer libros.
b. ÉL HACER LIBRO LEER.
c. Ella lee un libro.

(217) a. Después, merendamos leche y un bocadillo.
b. DESPUÉS, COMER BOCADILLO LECHE VASO.

c. Después, comemos un bocadillo y un vaso de leche.

(218) a. Luego, mamá nos lleva al parque.

b. DESPUÉS, MAMÁ ACOMPAÑAR PARQUE.

c. Después, mamá nos acompaña al parque.

(219) a. Allí Ana ve a sus amigas y se ponen a jugar.

b. ALLÍ PARQUE AMIGO+ MUJER("Ana") JUGAR.

c. Allí Ana juega con sus amigos.

(220) a. Al final, se pone muy sucia de correr y tirarse por el suelo.

b. AL-FINAL, SUCIO MOTIVO CORRER; SUELO CLS: 2 "persona revolcándose en el suelo" SUCIO ROPA.

(221) a. Al final, se ensucia porque corre y se tira por el suelo y se ensucia la ropa.

b. Luego, volvemos a casa.

c. FIN, IR CASA.

d. Luego, vamos a casa.

(222) a. Lo primero que hacemos es bañarnos.

b. PRIMERO QUÉ-HACER-? DUCHARSE LAVARSE.

c. Primero nos lavamos.

(223) a. Después, preparamos la ropa y la mochila para el día siguiente.

b. DESPUÉS, ROPA, MOCHILA POR-LA-MAÑANA ESE PREPARAR+.

c. Después, preparamos la ropa y la mochila de por la mañana.

(224) a. Mientras, mamá y papá nos hacen la cena.

b. ESE MIENTRAS PADRE-MADRE QUÉ-HACER-? POR-LA-NOCHE COMER.

c. Mientras, papá y mamá hacen la cena.

(225) a. Luego, nos vamos a la cama. Ana y yo dormimos en la misma habitación, pero cada una tiene su cama.

b. FIN, IR CAMA LOS-DOS HABITACIÓN, PRIVADO CAMA CLP: 5 5 "dos camas separadas".

c. Luego, vamos a la cama, a la misma habitación pero a camas separadas.

(226) a. Papá viene a leernos algún cuento hasta que nos dormimos.

b. PAPÁ CUENTO LEER HASTA LOS-DOS QUEDARSE-DORMIDO.

c. Papá lee un cuento hasta que nos quedamos dormidas.

(227) a. Algunas veces, de madrugada, Ana se despierta porque tiene sed.

b. ALGUNAS-VECES DE-MADRUGADA MUJER("Ana") DESPERTAR MOTIVO SED.

c. Algunas veces, de madrugada, Ana se despierta porque tiene sed.

(228) a. Entonces, llama a mamá o a papá y ellos le traen un vaso de agua.

b. ÉL QUÉ-HACER-?, PAPÁ O MAMÁ LLAMAR PARA AGUA VASO.

c. Llama a papá o a mamá para que le traigan un vaso de agua.

(229) a. Después sigue durmiendo hasta las 8.30 de la mañana.

b. DESPUÉS, QUEDARSE-DORMIDO HASTA HORA OCHO Y-MEDIA POR-LA-MAÑANA.

c. Después, se queda dormida hasta las ocho y media de la mañana.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford Studies in Typology and Linguistic Theory. Oxford University Press, New York, USA.

Alonso Baixeras, Pilar, Francisco Martínez Sánchez, Rubén Nogueira Fos, Juan Antonio Pinto Muñoz, M. Jesús Serna Serna and Ana M. Vázquez Aznar. 1998. *Signar: Enseñanza de la lengua de signos española*. CNSE.

Amate García, Mar. 2001. Personas sordas e intérpretes de LSE en España. *Faro del Silencio*, 182:16–21.

Aronoff, Mark, Irit Meir, Carol Padden and Wendy Sandler. 2004. Morphological universals and the sign language type. In G. Booj & J. van Marle (ed.), *Yearbook of morphology 2004*. Kluwer Academic Publishers, Dordrecht/Boston, pages 19–39.

Atserias, Jordi, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluís Padró and Muntsa Padró. 2006. FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 48–55, Genoa, Italy, May.

Baldassarri, Sandra and Francisco J. Royo-Santas. 2009. An automatic rule-based translation system to Spanish Sign Language. *New Trends on Human-Computer Interaction*, pages 1–11.

Bangham, J. Andrew, S. J. Cox, Richard Elliott, John R. W. Glauert, Ian Marshall, Sanja Rankov and Mark Wells. 2000. Virtual signing: capture, animation, storage and transmission - an overview of the ViSiCAST project. In *IEE Seminar on Speech and Language Processing for Disabled and Elderly People*, pages 1–7, London, England, UK, April.

Beesley, Kenneth R. and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Publications, Palo Alto, CA, USA.

Bertoldi, Nicola, Gabriele Tiotto, Paolo Prinetto, Elio Piccolo, Fabrizio Nunnari, Vincenzo Lombardo, Alessandro Mazzei, Rossana Damiano, Leonardo Lesmo and Andrea Del Principe. 2010. On the creation and the annotation of a large-scale Italian-LIS parallel corpus. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Languages Technologies*, pages 19–22, Valletta, Malta, May.

Braffort, Annelies, Annick Choisier, Christophe Collet, Patrice Dalle, Frédérick Gianni, Boris Lenseigne and Jérémie Segouat. 2004. Toward an annotation software for video of sign language, including image processing tools and signing space modelling. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pages 201–203, Lisbon, Portugal, May.

Brentari, Diane. 2010. *Sign Languages*. Cambridge University Press, New York, USA.

Bungeroth, Jan, Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, Andy Way and Lynette van Zijl. 2008. The ATIS sign language corpus. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*, pages 2943–2946, Marrakech, Morocco, May.

Bungeroth, Jan, Daniel Stein, Morteza Zahedi and Hermann Ney. 2006. A German Sign Language corpus of the domain weather report. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 2000–2003, Genoa, Italy, May.

Casacuberta, Francisco and Enrique Vidal. 2004. Machine translation with inferred stochastic finite-state transducers. *Computational Linguistics*, 30(2):205–225, June.

Chiu, Yu-Hsien, Chung-Hsien Wu, Hung-Yu Su and Chih-Jen Cheng. 2007. Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):28–39, January.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. Harper and Row.

Dangsaart, Srisavakon, Kanlaya Naruedomkul, Nick Cercone and Booncharoen Sirinaovakul. 2008. Intelligent Thai text - Thai sign translation for language learning. *Computers and Education*, 51(3):1125–1141, November.

Elliott, Ralph, John R. W. Glauert, Richard Kennaway, Ian Marshall and Éva Sáfár. 2008. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391.

Elliott, Ralph, Thomas Hanke, Ian Marshall, Constanze Schmaling, Éva Sáfár and Margriet Verlinden. 2001. Proto-text-to-sign notation. ViSiCAST Deliverable D5-3. Technical report.

Engberg-Pedersen, Elisabeth. 1989. Proformes en morphologie, syntaxe et discourse. In S. Quertinmont (ed.), *Études Européenes en Langue de Signes*. Edirsa, Brussels, Belgium, pages 35–52.

Engberg-Pedersen, Elisabeth. 1993. *Space in Danish Sign Language. The Semantics and Morphosyntax of the Use of Space in a Visual Language*, volume 19 of *International Studies on Sign Language Research and Communication of the Deaf*. Signum Press, Hamburg, Germany.

Fernández-Soneira, Ana M.ª. 2008. *La cantidad a manos llenas: La expresión de la cuantificación en la lengua de signos española*. Investigaciones sobre la lengua de signos española y la comunidad sorda. Fundación CNSE.

Forcada, Mikel L., Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O'Regan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez and Francis M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144.

Fundación CNSE. 2008. *Diccionario normativo de la lengua de signos española*. Fundación CNSE.

Greenberg, Joseph H. 1963. *Universals of Languages*. MIT Press, Cambrigde, Massachusets, USA.

Gutiérrez-Sigut, E., B. Costello, C. Baus and M. Carreiras-Valiña. 2012. LSE-Sign: una herramienta para la selección de signos de la lengua de signos española. http://www.bcbl.eu/databases/lse.

Gutiérrez-Sigut, Eva and Manuel Carreiras-Valiña. 2009. *El papel de los parámetros fonológicos en el procesamiento de la lengua de signos española*. Investigaciones sobre la lengua de signos española y la comunidad sorda. Fundación CNSE.

Gómez-Alcaraz, Encarnación. 2011. Analizador léxico del sistema de escritura alfabética de lengua de signos española. Master's thesis, Escuela Politécnica Superior – Universidad de Alicante, December.

Hanke, Thomas. 2002. iLex - A tool for sign language lexicography and corpus analysis. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 923–926, Las Palmas, Spain, May.

Hanke, Thomas, G. Langer and C. Metzger, 2001. *Interface Definitions*, chapter Encoding non-manual aspects of sign language. ViSiCAST Deliverable D5-1.

Hengeveld, Kees, Jan Rijkhoff and Anna Siewierska. 2004. Parts of speech systems and word order. *Journal of Linguistics*, 40:527–570, November.

Herrero Blanco, Ángel. 2003. *Escritura alfabética de la lengua de signos española: Once lecciones*. Publicaciones de la Universidad de Alicante.

Herrero Blanco, Ángel. 2009. *Gramática didáctica de la lengua de signos española (LSE)*. Ediciones SM, 1 edition.

Herrero Blanco, Ángel, Juan J. Alfaro and Inmaculada Cascales, 2001. *Lecciones de escritura alfabética de la LSE*. Universidad de Alicante.

Herrero Blanco, Ángel and Ana Peidro. 2007. Los tipos de plural en la lengua de signos española. In *Actas del IV Congreso de Lingüística General*, pages 1051–1064.

Hockett, Charles F. 1960. The origin fo speech. *Scientific American*, 203:88–96.

Huenerfauth, Matt. 2003. A survey and critique of American Sign Language natural language generation and machine translation systems. Technical report, University of Pennsylvania, September.

Huenerfauth, Matt. 2004. Spatial representation of classifier predicates for machine translation into American Sign Language. In *Workshop on the Representation and Processing of Signed Languages, 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.

Huenerfauth, Matt. 2006. *Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation*. Ph.D. thesis, University of Pennsylvania, January.

INE, 1999. *Encuesta sobre discapacidades, deficiencias y estado de salud*. Instituto Nacional de Estadística.

INE, 2003. *Las personas con discapacidad y su relación con el empleo. Encuesta de población activa del segundo trimestre de 2002*. Instituto Nacional de Estadística.

INE, 2008. *Encuesta sobre discapacidades, autonomía personal y situaciones de dependencia*. Instituto Nacional de Estadística.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association – A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, Cambridge, USA.

Isabelle, Pierre and George Foster. 2005. Machine Translation Part I - Overview. *Encyclopedia of Language and Linguistics*, 15:404–422.

Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description.* Mouton, The Hague, The Netherlands.

Johnston, Trevor. 1991a. Spatial syntax and spatial in the inflection of signs for marking person and location on Auslan. *International Journal of Sign Linguistics*, 2(1):29–62.

Johnston, Trevor. 1991b. Transcription and glossing of sign language texts: Examples from Auslan (Australian Sign Language). *International Journal of Sign Linguistics*, 2(1):3–28.

Kanis, Jakub, Jiri Zahradil, Filip Jourcicek and Lůdek Müller. 2007. Czech-Sign speech corpus for semantic based machine translation. *Lecture Notes in Computer Science*, 4188:613–620.

Kaplan, Ronald M. and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, 20(3):331–378.

Kennaway, Richard, John R. W. Glauert and Inge Zwitserlood. 2007. Providing signed content on the internet by synthesized animation. *ACM Transactions on Compututer-Human Interaction*, 14(3).

Kipp, M. 2001. ANVIL - A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (EuroSpeech-2001)*, pages 1367–1370, Aalborg, Denmark, September.

Kiss, Katalin É. (ed.). 1995. *Discourse Configurational Languages.* Oxford University Press.

Klima, Edward S. and Ursula Bellugi. 1979. *The signs of language.* Harvard University Press, Cambridge, MA, USA.

Knowles, Gerry and Zuraidah Mohd Don. 2004. The notion of a "lemma": Headwords, roots and lexical sets. *International Journal of Corpus Linguistics*, 9(1):69–81.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of Demonstration and Poster Sessions at the 45th Annual Meeting of the Association for Computational Linguistics (ACL-2007)*, pages 177–180, Prague, Czech Republic, June.

Lang, Mervyn Francis. 1990. *Spanish Word Formation: Productive Derivational Morphology in the Modern Lexis*. Routledge, London & New York.

Leeson, Lorraine and John Saeed. 2012. Word order. In Ronald Pfau, Markus Steinbach, and Bencie Woll (eds.), *Sign Language: An International Handbook*. De Gruyter Mouton, pages 245–265.

Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10:707–710.

Li, Charles N. and Sandra A. Thompson. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and Topic*. Academic Press, London/New York, pages 457–461.

Liddell, Scott and Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies*, 64:192–277.

López-Colino, Fernando J. and José Colás. 2011. The synthesis of LSE classifiers: From representation to evaluation. *Journal of Universal Computer Science*, 17(3):399–425, February.

López-Ludeña, Verónica, Rubén San-Segundo, Ricardo de Córdoba, Javier Ferreiros, Juan Manuel Montero and José Manuel Pardo. 2011a. Factored translation models for improving a speech into Sign Language translation system. In *Proceedings of the 12th Annual Conference of the International Speech Communication (InterSpeech-2011)*, pages 1605–1608, Makuhari, Chiba, Japan, September.

López-Ludeña, Verónica, Rubén San-Segundo, S. Lutfi, J.M. Lucas-Cuesta, J.D. Echevarry and B. Martínez-González. 2011b. Source language categorization for improving a speech into sign language translation system. In *Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2011)*, pages 84–93, Edinburgh, Scotland, UK, July.

López-Ludeña, Verónica, Rubén San-Segundo, Juan Manuel Montero, Ricardo de Córdoba, Javier Ferreiros and José Manuel Pardo. 2012. Automatic categorization for improving Spanish into Spanish Sign Language machine translation. *Computer Speech & Language*, 26(3):149–167.

Marimon, Montserrat. 2010. The Tibidabo Treebank. *Procesamiento del Lenguaje Natural*, 45:113–119.

Marimon, Montserrat, Beatriz Fisas, Núria Bel, Blanca Arias, Silvia Vázquez, Jorge Vival di, Sergi Torner, Marta Villegas and Mercé Lorente. 2012. The IULA Treebank. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC-2012)*, pages 1920–1926, Istanbul, Turkey.

Marshall, Ian and Éva Sáfár. 2004. Sign language generation in an ALE HPSG. In Stefan Müller (ed.), *Proceedings of the HPSG-2004 Conference*, pages 189–201, Center for Computational Linguistics, Katholieke Universiteit Leuven. CSLI Publications.

Massó, Guillem and Toni Badia. 2010. Dealing with sign language morphemes in statistical machine translation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Languages Technologies*, pages 154–157, Valletta, Malta, May.

McBurney, Susan L. 2002. Pronominal reference in signed and spoken languages. In K. A. Cormier & D.G. Quinto Meier, R. P. (ed.), *Modality and Structure in Signed and Spoken Languages*. Cambridge University Press, Cambridge, USA, pages 329–69.

MECD, 2000. *Serie informes - Centro Nacional de Información y Comunicación Educativa (CNICE)*. Ministerio de Educación, Cultura y Deporte.

Minguet Soto, Amparo (ed.). 2000. *Signolingüística. Introducción a la lingüística de la LSE*. Fundación FESORD, Valencia, Spain.

Morrissey, Sara. 2008. *Data-Driven Machine Translation for Sign Languages*. Ph.D. thesis, School of Computing. Dublin City University, August.

Morrissey, Sara. 2011. Assessing three representation methods for Sign Language machine translation and evaluation. In *Proceedings of the 15th Annual Meeting of the European Association for Machine Translation (EAMT-2011)*, pages 137–144, Leuven, Belgium.

Morrissey, Sara and Andy Way. 2006. Lost in translation: the problems of using mainstream MT evaluation metrics for Sign Language translation. In *Proceedings of the 5th SALTMIL Workshop on Minority Languages at LREC-2006*, pages 91–98, Genoa, Italy, May.

Neidle, Carol. 2002. Signstream: A database tool for research on visual-gestural language. *Sign Language and Linguistics*, 4(1/2):203–214.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13:95–135, 5.

Nöth, Winfried. 1990. *Handbook of semiotics*. Advances in Semiotics. Indiana University Press, Indiana, USA.

Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, March.

Padden, Carol. 1990. *Sign language research: Theoretical issues*. Gallaudet University Press, Washington DC, USA.

Papinemi, Kishore, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, pages 311–318, Philadelphia, Pennsylvania, USA, July.

Parkhurst, Steven and Dianne Parkhurst. 2001. *Un estudio lingüístico: Variación de las lenguas de signos en España*. Revista Española de Lingüística de Lengua de Signos (RELLS), Madrid, Spain.

Pfau, Ronald, Markus Steinbach and Bencie Woll (eds.). 2012. *Sign Language: An International Handbook*. De Gruyter Mouton.

Pollard, Carl J. and Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. University of Chicago Press, Chicago, USA.

Prillwitz, Siegmund, Regina Leven, Heiko Zienert, Thomas Hanke and Jan Henning. 1989. HamNoSys Version 2.0: Hamburg Notation System for Sign Language: An introductory guide. *International Studies on Sign Language and the Communication of the Deaf*, 5.

Real Academia Española. 2001. *Diccionario de la lengua española*. Espasa, Madrid, Spain, 22 edition.

Real Academia Española. 2011. *Nueva gramática básica de la lengua española*. Espasa.

Rijkhoff, Jan and Eva van Lier (eds.). 2013. *Flexible Word Classes: Typological studies of underspecified parts of speech*. Oxford Academic Press.

Roark, Brian and Richard W. Sproat. 2007. *Computational Approaches to Morphology and Syntax*. Oxford Surveys in Syntax & Morphology. OUP Oxford.

Rodríguez González, M. Ángeles. 1992. *Lenguaje de signos*. Fundación CNSE, Barcelona, Spain.

Rodríguez Ortiz, Isabel de los Reyes. 2005. *Comunicar a través del silencio: Las posibilidades de la lengua de signos española*. Universidad de Sevilla, April.

Rubino, Francesco, Alan Hayhurst and Josif Guejlman. 1975. *Gestuno. International sign language of the deaf*. British Deaf Association, Carlisle, UK.

Sáfár, Éva and Ian Marshall. 2002. Sign language translation via DRT and HPSG. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pages 58–68, México.

San-Segundo, Rubén, Verónica López-Ludeña, R. Martín, D. Sánchez and A. García. 2010. Language resources for Spanish - Spanish Sign Language (LSE) translation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Languages Technologies*, pages 208–211, Valletta, Malta, May.

San-Segundo, Rubén, Juan Manuel Montero, Ricardo de Córdoba, V. Sama, F. Fernández, Luis D'Haro, Verónica López-Ludeña, D. Sánchez and A. García. 2011. Design, development and field evaluation of a Spanish into Sign Language translation system. *Pattern Analysis & Applications*, pages 1–22.

Sandler, Wendy. 1989. *Phonological representation of the sign: linearity and non-linearity in American Sign Language*. Floris, Dordrecht, The Netherlands.

Sandler, Wendy and Diane Lillo-Martin. 2006. *Sign Language and Linguistic Universals*. Cambridge University Press, New York, USA.

Santana-Suárez, Octavio, Francisco Javier Carreras-Riudavets and José Rafael Pérez-Aguiar. 2004. *Relaciones morfoléxicas sufijales para el procesamiento del lenguaje natural*. Editorial MILETO, Madrid, Spain.

Santana-Suárez, Octavio, Francisco Javier Carreras-Riudavets and José Rafael Pérez-Aguiar. 2005. *Relaciones morfoléxicas prefijales para el procesamiento del lenguaje natural*. Editorial MILETO, Madrid, Spain.

Santana-Suárez, Octavio, Francisco Javier Carreras-Riudavets and José Rafael Pérez-Aguiar. 2006. *Relaciones morfoléxicas parasintéticas para el procesamiento del lenguaje natural*. Editorial MILETO, Madrid, Spain.

Schembri, Adam. 2003. 1: Rethinking 'classifiers' in Signed Languages. In *Perspectives on classifier constructions in sign languages*. Psychology Press, pages 3–34.

Schmaling, Constanze and Thomas Hanke. 2001. HamNoSys 4.0. Technical report.

Schreurs, Linda. 2006. The distinction between formally and semantically related noun-verb pairs in Sign Language of the Netherlands (NGT). Master's thesis, Universiteit van Amsterdam.

Schulmeister, Rolf. 2001. The ViSiCAST Project: Translation into sign language generation of sign language by virtual humans (avatars) in television, WWW and face-to-face transactions. In Constantine Stephanidis (ed.), *Universal Access in HCI: Towards an Information Society for All*, volume 3. Erlbaum, Hillsdale, NJ, USA, pages 431–435.

Shield, Aaron and Jason Baldridge. 2009. A morphological analyzer for verbal aspect in American Sign Language. In Nicholas Gaylord, Stephen Hilderbrand, Heeyoung Lyu, Alexis Palmer, and Elias Ponvert (eds.), *Computational Linguistics for Less-Studied Languages*, volume 10. CSLI Publications, pages 125–138.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of Association for Machine Translation in the Americas*, pages 223–231.

Speers, d'Armond L. 2001. *Representation of American Sign Language for Machine Translation*. Ph.D. thesis, Georgetown University, December.

Stein, Daniel, Christoph Schmidt and Hermann Ney. 2010. Sign language machine translation overkill. In Marcello Federico, Ian Lane, Michael Paul, and François Yvon (eds.), *International Workshop on Spoken Language Translation*, pages 337–344, Paris, France, December.

Stokoe, William C. 1960. Sign language structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics: Occasional Papers*, 8:1–78.

Stroppa, Nicolas and Andy Way. 2006. MATREX: DCU machine translation system for IWSLT 2006. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT-2006)*, pages 31–36, Kyoto, Japan, November.

Supalla, Ted. 1986. The classifier system in American Sign Language. In Colette Craig (ed.), *Noun Classes and Categorization*, number 7 in Typological Studies in Language. John Benjamins, pages 181–214.

Supalla, Ted and Elissa L. Newport. 1978. How many seats in a chair? The derivation of nouns and verbs in ASL. In P. Siple (ed.), *Understanding language through sign language research*. Academic Press, New York, USA, pages 91–132.

Sutton, Valerie J. 1973. *Sutton movement shorthand, the classical ballet key, key one.* The Sutton Movement Writing Press, December.

Sutton, Valerie J. 1981. *Signwritting for everyday use.* The Sutton Movement Writting Press, Newport Beach, CA, USA.

Sutton-Spence, Rachel and Bencie Woll. 1999. *The linguistics of British Sign Language: An introduction.* Cambridge University Press, Cambridge, UK.

Tai, Terry, Wojciech Skut and Richard Sproat. 2011. Thrax: An open source grammar compiler built on OpenFst. In *ASRU 2011*, Waikoloa Resort, Hawaii, USA, December.

Varga, Dániel, Peter Halácsy, András Kornai, Victor Nagy, László Németh and Viktor Trón. 2007. Parallel corpora for medium density languages. In G. Angelova N. Nicolov, K. Bontcheva and R. Mitkov (eds.), *Recent Advances in Natural Language Processing IV. Selected papers from RANLP-2005.* John Benjamins, pages 247–258.

Vauquois, Bernard. 1968. A survey of formal grammars and algorithms for recognition and transformation in machine translation. In *Proceedings of 6th the IFIP Congress*, pages 254–260.

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

Veale, Tony and Alan Conway. 1994. Cross modal comprehension in ZARDOZ: An English to Sign-Language translation system. In *Proceedings of the 7th International Workshop on Natural Language Generation (INLG-1994)*, pages 249–252, Kennebunkport, Maine, USA, June.

Villameriel, Saúl. 2008. *Marcadores del discurso en la lengua de signos española y en el español oral: Un estudio comparativo.* Ph.D. thesis, Universidad de León, June.

Vossen, Piek (ed.). 1998. *EuroWordNet. A multilingual database with lexical semantic networks.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wintner, Shuly. 2007. Finite-state technology as a programming environment. In A. Gelbukh (ed.), *Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing-2007)*, volume 4394 of *Lecture Notes in Computer Science*, pages 97–106, Berlin and Heidelberg, Germany. Springer.

Wittenburg, Peter, Stephen C. Levinson, Sotaro Kita and Hennie Brugman. 2002. Multimodal annotations in gesture and sign language studies. In *Proceedings of the 3rd*

International Conference on Language Resources and Evaluation (LREC-2002), pages 176–182, Las Palmas, Spain, May.

Zeshan, Ulrike. 2007. Roots, leaves and branches – the typology of sign languages. In R.M. de Quadros (ed.), *Sign languages: Spinning and unraveling the past, present and future*. Editora Arara Azul, Petrópolis, Brazil, pages 671–695.

Zhao, Liwei, Karin Kipper, William Schuler, Christian Vogler, Norman I. Badler and Martha Palmer. 2000. A machine translation system from English to American Sign Language. In *Proceedings of the 4th Conference of the Association for Machine Translation in the Americas on Envisioning Machine Translation in the Information Future (AMTA-2000)*, pages 191–193, Cuernavaca, México, October.