Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 28 (2014) 788–811

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

A rule-based translation from written Spanish to Spanish Sign Language glosses☆

Jordi Porta*, Fernando López-Colino, Javier Tejedor, José Colás

Human Computer Technology Laboratory, Escuela Politécnica Superior – Universidad Autónoma de Madrid, Av. Francisco Tomás y Valiente, 11, 28049 Madrid, Spain

Received 1 February 2012; received in revised form 26 August 2013; accepted 11 October 2013
Available online 27 October 2013

Abstract

One of the aims of Assistive Technologies is to help people with disabilities to communicate with others and to provide means of access to information. As an aid to Deaf people, we present in this work a production-quality rule-based machine system for translating from Spanish to Spanish Sign Language (LSE) glosses, which is a necessary precursor to building a full machine translation system that eventually produces animation output. The system implements a transfer-based architecture from the syntactic functions of dependency analyses. A sketch of LSE is also presented. Several topics regarding translation to sign languages are addressed: the lexical gap, the bootstrapping of a bilingual lexicon, the generation of word order for topic-oriented languages, and the treatment of classifier predicates and classifier names. The system has been evaluated with an open-domain testbed, reporting a 0.30 BLEU (BiLingual Evaluation Understudy) and 42% TER (Translation Error Rate). These results show consistent improvements over a statistical machine translation baseline, and some improvements over the same system preserving the word order in the source sentence. Finally, the linguistic analysis of errors has identified some differences due to a certain degree of structural variation in LSE.

Keywords: Machine translation; Spanish; Spanish Sign Language; LSE; Deaf people communication

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, since translation helps

☆ This paper has been recommended for acceptance by R.K. Moore.

* Corresponding author. Tel.: +34 91 497 33 64; fax: +34 91 497 22 35.
E-mail addresses: jordi.porta@uam.es (J. Porta), fj.lopez@uam.es (F. López-Colino), javier.tejedor@uam.es (J. Tejedor), jose.colas@uam.es (J. Colás).

0885-2308/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.csl.2013.10.003

Deaf and Hearing communities to communicate with each other, and provides Deaf people with the same opportunities to access information as everyone else.

1.1. Sign Languages

Sign Languages (SLs) exploit a different physical medium from the oral-aural system of spoken languages. SLs are gestural-visual languages, and this difference in modality causes SLs to constitute another branch within the typology of languages. However, there are still many myths around SLs. One of the most common and enduring myths is that sign language is universal; however, every Deaf community has its own SL, even within the same country. For example, in Spain, apart from Spanish Sign Language (LSE), there exists another recognised SL, 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 SL families. American Sign Language (ASL) and British Sign Language (BSL), however, despite the fact that both are SLs 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 SL is extended, dialects and varieties are developed. In other words, SLs are natural languages that arise spontaneously in Deaf communities to fulfil the function of communication.

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 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 SLs as "real" human languages have been made since the work of W. Stokoe on the ASL (Stokoe, 1960). SLs in developed countries, especially SLs 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 Zesman (2007), the state of knowledge regarding SLs has developed like a mosaic with many untiled gaps, but this is increasingly giving typologists a clearer picture of the range of diversity in SLs. Some cross-linguistic and typological studies of SLs, such as those by Sandler and Lillo-Martin (2006) or Brentari (2010), have shed light on both the universals and the diversity of SLs, contributing to the understanding of human languages in general.

The lack of a writing system is characteristic of SLs, and one shared with two-thirds of the spoken languages of the world. Strictly speaking, the only way of representing SLs is to use motion pictures. However, several notational systems exist. The most important today are SignWriting (Sutton, 1974) and HamNoSys (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 SLs, with the same objective as the International Phonetic Alphabet (IPA) for spoken languages. There is another alphabetic writing system, designed specifically for LSE, and called SEA (*Sistema de escritura alfabética*) (Herrero et al., 2001); this uses the Latin alphabet, and has LSE's phonology as its basis. As this paper focuses primarily on syntax, glosses will be used here instead of a phonological notation. 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 (MT) system needs to produce an animation to be considered a complete and useful system.

Most contemporary works on SLs 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 SLs, but also to the study of all other descriptive levels found in spoken languages. However, from the point of view of natural language processing, SLs 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 SLs, since it in fact applies to most of the languages of the world.