Automatic Translation from German to Synthesized Swiss German Sign Language

Thesis

presented to the Faculty of Arts and Social Sciences

of the University of Zurich

for the degree of Doctor of Philosophy

by Sarah Ebling

Accepted in the spring semester 2016

on the recommendation of the doctoral committee:

Prof. Dr. Martin Volk (main advisor)

Prof. Dr. Pierrette Bouillon

Prof. Dr. Elvira Glaser

Zurich, 2016

Abstract

This thesis presents research in automatic translation from German to synthesized Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS), the sign language of the German-speaking area of Switzerland. The research is connected to the use case of building a system that translates written German train announcements of the Swiss Federal Railways into DSGS, the ultimate output consisting of a signing avatar. The thesis centers around three areas: corpus linguistics, sign language machine translation, and sign language animation.

Being the first work to apply machine translation and animation to DSGS, the thesis establishes the prerequisites of successful automatic processing of this language. Moreover, it specifies the information necessary to bridge the gap between sign language machine translation and sign language animation. The thesis then reports on experiments in sign language machine translation using statistical methods. It also presents a solution for automatically generating non-manual information, i.e., information pertaining to components other than the hands.

The thesis further makes a contribution to quality improvement of signing avatar systems. The modifications made to an avatar system are described, and the results of an evaluation of the resulting DSGS avatar are presented. Lastly, the thesis reports on work in synthesizing the finger alphabet of DSGS.

Abstract

Diese Arbeit präsentiert Forschung zu automatischer Übersetzung von Deutsch in synthetisierte Deutschschweizerische Gebärdensprache (DSGS), die Gebärdensprache des deutschsprachigen Teils der Schweiz. Die Forschung steht im Zusammenhang mit einem System, das deutschsprachige Zugansagen der Schweizerischen Bundesbahnen in schriftlicher Form in die DSGS übersetzt, wobei die finale Ausgabe aus einem gebärdenden Avatar besteht. Die Arbeit ist in drei Bereiche aufgeteilt: Korpuslinguistik, Gebärdensprachübersetzung und Gebärdensprachanimation.

Als erste Arbeit, die sich mit der maschinellen Übersetzung nach und der Synthetisierung von DSGS befasst, legt sie die Voraussetzungen für eine automatische Verarbeitung dieser Sprache dar. Sie spezifiziert zudem die Art der Information, die notwendig ist, um Gebärdensprachübersetzung und Gebärdensprachanimation miteinander zu kombinieren. Die Arbeit stellt sodann Experimente in Gebärdensprachübersetzung unter Anwendung statistischer Methoden vor. Sie präsentiert auch einen Ansatz zur automatischen Generierung nicht-manueller Information, d.h. von Information, die nicht die Hände betrifft.

Die Arbeit leistet auch einen Beitrag zur Verbesserung der Qualität von Gebärdensprachavataren. Es werden Modifikationen an einem bestehenden Avatarsystem beschrieben und die Resultate einer Evaluation des resultierenden DSGS-Avatars vorgestellt. Zum Abschluss präsentiert die Arbeit einen Ansatz zur Synthetisierung des Fingeralphabets der DSGS.

Acknowledgements

First and foremost, my thanks go to my advisor Martin Volk, who has given me the opportunity to work on the topic of my choice for my PhD, convinced that intrinsic motivation is the best driving force for such an endeavour. He followed my work with great interest and was always ready to provide valuable input. His quick turnaround and his humor are much appreciated. I would also like to thank the two other members of my doctoral committee, Pierrette Bouillon and Elvira Glaser, for taking the time to review my thesis and providing valuable feedback.

I had first made contact with the topic of automatic sign language processing at Dublin City University through the work of Sara Tucker (née Morrissey) and Robert Smith.

Carrying out research on automatic sign language processing in Switzerland would not have been possible without the extensive support of Penny Boyes Braem, who not only generously shared the resources she had built over the past thirty years with me but also to this day provides feedback and input to many persons interested in DSGS research, including myself. Penny's expertise is unmatched and her passion deeply inspiring. My heartfelt thanks go to her for continuing to dedicate her time to the advancement of sign language research in Switzerland.

I am very grateful to Katja Tissi and Sandra Sidler-Miserez for their work, their reliability, and, more generally, their openness to the topic of automatic sign language processing. I would like to thank the Federal Bureau for the Equality of People with Disabilities (grant no. 12.I.016) and the Max Bircher Foundation for funding their work. The Swiss National Science Foundation allowed me to conduct parts of my research in the United States as a Doc.Mobility fellowship (grant no. 155 263).

I am greatly indebted to John Glauert for always responding to questions regarding the JASigning software quickly and thoroughly. Ralph Elliott, Richard Kennaway, and Vince Jennings also offered valuable insights into JASigning at various stages of my work.

I would like to thank Michael Böhm of the Swiss Federal Railways for providing the German train announcements. My thanks also go to the Swiss Deaf Association for giving me access to their video recording studio and providing additional support in the person of Toni Koller. I am also grateful to the participants of the focus group study aimed at evaluating the DSGS avatar signing train announcements. Students of the DSGS interpreting programme at the University of Applied Sciences of Special Needs Education Zurich collected signs for places with train stations in Switzerland. I am also indebted to the participants of the donate-a-sign call and the online study aimed at evaluating the comprehensibility of synthesized DSGS fingerspelling sequences.

Thomas Hanke has been generous in sharing his expertise in the field of automatic sign language processing on various occasions in addition to providing support for the iLex software.

My thanks also go to the people from the two institutions in the United States where I spent time: Rosalee Wolfe, John McDonald, Robyn Moncrief, and Souad Baowidan went out of their way to make my stay in Chicago a pleasant one. Matt Huenerfauth shared his extensive knowledge in the areas of computational linguistics, sign language, and human-computer interaction with me.

I am also grateful for the support of my former and present colleagues at the Institute of Computational Linguistics, especially Rico Sennrich, Anne Göhring, Alexandra Bünzli, Simon Clematide, and Manfred Klenner.

Lastly, I would like to extend my deepest gratitude to my parents and sister and to my partner Tobias.

Contents

List of figures

List of tables

List of abbreviations

ASL American Sign Language

BLEU Bilingual Evaluation Understudy [Metric]

BOMP Bonn Machine-Readable Pronunciation Dictionary

BSL British Sign Language

CRF Conditional Random Fields

CSL Chinese Sign Language

DGS German Sign Language (*Deutsche Gebärdensprache*)

DSGS Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*)

DTD document type definition

HamNoSys Hamburg Notation System for Sign Languages

IS International Sign

ISL Irish Sign Language

LSF French Sign Language (*Langue des Signes Française*)

LSF-CH Swiss French Sign Language (*Langue des Signes Française Suisse*)

NGT Sign Language of the Netherlands (*Nederlandse Gebarentaal*)

NIST National Institute of Standards and Technology [Metric]

ÖGS Austrian Sign Language (*Österreichische Gebärdensprache*)

PER Position-Independent Word Error Rate

SAMPA Speech Assessment Methods Phonetic Alphabet

SBB Swiss Federal Railways (*Schweizerische Bundesbahnen*)

SiGML Signing Gesture Markup Language

SMT statistical machine translation

TSL Taiwanese Sign Language

TER Translation Edit Rate

WER Word Error Rate

Abbreviations are expanded at the beginning of each chapter.

Chapter 1

Introduction

Our daily lives abound with situations in which we rely on information being provided to us, be it in the car, at the bus or train station, at work, in school, or at home. For certain groups of people, access to information is prohibitively limited. This is true, for example, for Deaf[1] persons who rely on information being conveyed to them in sign language.

In discussions of the necessity of providing information in sign language, often no mention is made of the average reading and writing level of a Deaf adult in a surrounding spoken language[2] to correspond to that of a hearing 10 year-old child (Gutjahr, 2006; Traxler, 2000; Wauters, 2005). Several reasons have been given for this (Konrad, 2011): Firstly, precisely because of their hearing loss, Deaf persons do not have access to the phonological basis of a spoken language. Moreover, a Deaf person lacks auditory feedback when reading out loud. Linguistic differences between sign languages and spoken languages are a further reason why acquiring a spoken language is difficult for Deaf signers. Sign languages also do not as of yet have a widely accepted writing system that would promote the concept of literacy. Lastly, spoken languages may carry a negative connotation for some sign language users due to negative experiences they made in schools in which they were taught orally.

[1] It is a widely recognized convention to use the upper-cased word *Deaf* for describing members of the linguistic community of sign language users and, in contrast, to use the lower-cased word *deaf* when describing the audiological state of a hearing loss (Morgan & Woll, 2002).

[2] A *spoken language* is a language that is not signed, whether it is represented as speech or text. Alternative terms are *vocal language* and *oral language*.

While the amount of information that is provided in sign language differs from country to country, it is generally significantly lower than that of information available in a surrounding spoken language. *Automatic sign language processing*, a sub-field of natural language processing (NLP), provides a way of reducing this imbalance. Automatic sign language processing comprises applications such as sign language recognition, sign language synthesis/animation, or sign language translation (Sáfár & Glauert, 2012). For each of these applications, important contributions have been made in the past decades, but the existing body of research is still considerably smaller than that of the field of automatic spoken language processing.

Specifically, while individual sign language processing applications have been developed, they have rarely been combined into a pipeline that would allow for fully automatic translation of spoken language into sign language, of sign language into spoken language, or of sign language into sign language. Figure 1.1 visualizes these three pipelines: Pipeline a) requires at the very least a machine translation step from spoken language into sign language followed by a sign language animation step. If the initial input is speech as opposed to text, the core pipeline is preceded by a speech recognition step. Pipeline b) requires a sign language recognition step and a subsequent machine translation step from sign language into spoken language. If the ultimate output is to be speech rather than text, a speech synthesis step ensues. Finally, pipeline c) requires a sign language recognition step, a machine translation step from one sign language into another, and a sign language animation step.

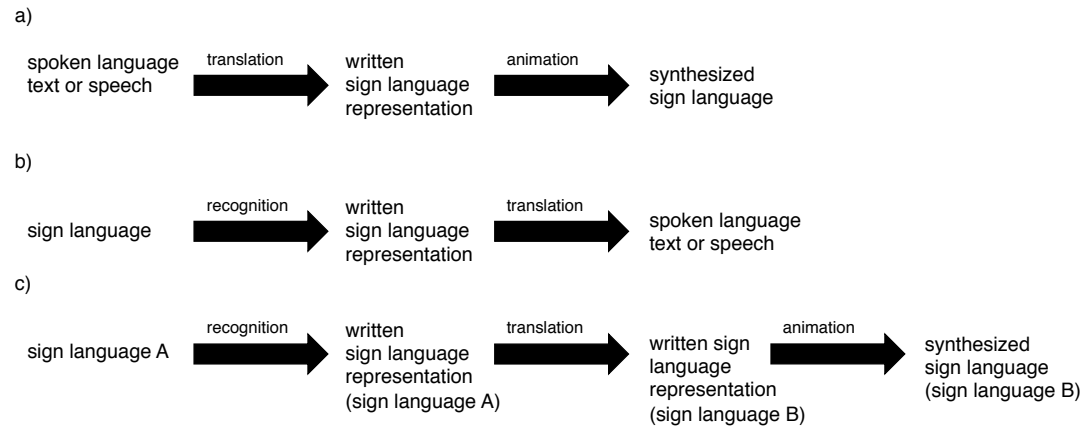


Figure 1.1: Automatic sign language processing: Three pipelines

The primary goal of any system instantiating any of these three pipelines cannot and should not be to replace human sign language interpreters. As Huenerfauth & Hanson (2009) state, "no computer system is capable of providing the same level of sophisticated and subtle translation

that a qualified professional interpreter can" (p. 12). Instead, systems of the aforementioned kinds should be applied to settings in which interpreters are not naturally available (e.g., rendering web content) or in which content is standardized as well as highly dynamic (e.g., rendering passenger information in the public transportation domain).

The research presented in this thesis stems from a project that established a pipeline of type a) (cf. Figure 1.1): A system was built that translates written German train announcements of the Swiss Federal Railways (*Schweizerische Bundesbahnen*) (SBB) into Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS), the final output being a virtual signer, or *signing avatar*, as shown in Figure 1.2. The research that informed the development of the system centered around three areas: (1) corpus linguistics, (2) machine translation, and (3) sign language animation.

Figure 1.2: Signing avatar

1.1 Research questions

The research questions guiding my work in these three areas were:

1. **Corpus linguistics:** What are the steps necessary to build a parallel corpus for use in statistical machine translation from German to DSGS and subsequent DSGS animation?
2. **Machine translation:**
   (a) What are the steps necessary to build a domain-specific machine translation system, i.e., a system that translates German train announcements into DSGS?
   (b) How well can such a system perform as measured by automatic evaluation metrics?

(c) How can non-manual information in signing (i.e., information pertaining to components other than the hands) be included in the output of such a system?

3. **Sign language animation:** What are the steps necessary to customize a signing avatar system so as to render it suitable for displaying synthesized DSGS?

As a result of addressing the above research questions, the present thesis makes several contributions to the field of automatic sign language processing:

- It represents the first work to apply machine translation and animation to DSGS. As such, it naturally contributes to an understanding of the prerequisites of automatically processing this language.
- It specifies, in a language-independent manner, the type of information to be included in a parallel corpus for use in sign language machine translation and subsequent sign language animation.
- It presents experiments in sign language machine translation using statistical methods.
- It represents one of few works to propose a solution for including non-manual information in the output of a sign language machine translation system. The proposed solution is language-independent.
- It presents work in improving the quality of a signing avatar system.

1.2 Overview of chapters

The thesis is structured as follows: Chapter 2 gives an introduction to those linguistic properties of sign languages that are relevant for understanding the challenges posed by automatic sign language processing. In particular, Section 2.1 introduces the articulators in sign languages; Section 2.2 presents ways of representing sign language in written form; Section 2.3 discusses constituent order in sign languages; Section 2.4 introduces the concept of iconicity, discusses its prevalence in signing, and describes a study undertaking a lexical comparison between German Sign Language (*Deutsche Gebärdensprache*) (DGS) and DSGS; Section 2.5 discusses fingerspelling in sign languages; and Section 2.6 presents communication systems similar to sign languages.

The remaining chapters are arranged such that each covers one of the focus areas of this thesis: sign language corpora, sign language machine translation, and sign language animation. Within each chapter, the first few sections present the current state of research. The remaining sections then give an account of my own contribution to the respective area.

In such a way, Chapter 3 discusses the process of compiling a sign language corpus, from obtaining raw data (Section 3.1) to creating primary data (Section 3.2), secondary data (Section 3.3), and metadata (Section 3.4). Section 3.5 gives examples of previous sign language corpora used in automatic sign language processing. Section 3.6 then covers the process of building a parallel corpus of German/DSGS train announcements for use in sign language machine translation and subsequent sign language animation.

Chapter 4 deals with sign language machine translation. The paradigms of machine translation (Section 4.1) and ways of automatically evaluating machine translation output (Section 4.2) are introduced and previous approaches to limited-domain statistical sign language machine translation described (Section 4.3). Section 4.4 introduces my own work in automatically translating written German train announcements into DSGS. As a separate sub-problem, Section 4.5 presents my solution for automatically generating non-manual information.

Chapter 5 is dedicated to sign language animation: Different approaches to animation (Section 5.1) and ways of evaluating animation systems (Section 5.2) are discussed. I then report on my own work in enhancing an animation system (Section 5.3). Lastly, I present my work in the field of automatic fingerspelling synthesis using a different animation system (Section 5.4).

Each chapter concludes with a summary section. Chapter 6 offers an overall conclusion as well as an outlook on future work.

Chapter 2

Sign languages

Contrary to popular belief, no universal sign language exists. Approximately 120 sign languages are known to date, and new ones are still being discovered (Zeshan, 2012). All of these sign languages are natural languages and, as such, fully developed linguistic systems with a grammar and a vocabulary (Johnston & Schembri, 2007).[1]

Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS) is the sign language of the German-speaking area of Switzerland. It is not among the official languages of Switzerland (those being German, French, Italian, and Romansh), for which the number of speakers is regularly determined through an official census. The exact statistics for DSGS users is therefore unknown; estimates range between 5 500 (Boyes Braem, 2012b) and 6 000 (Lewis, 2009) Deaf signers. For any sign language, approximately 5% of its Deaf users typically have Deaf parents and are considered *native signers*, while the remaining 95% are children of hearing parents and, hence, *non-native signers* (Mitchell & Karchmer, 2004). In Switzerland, in addition to the Deaf DSGS users mentioned above, an estimated 13 000 hearing persons acquire DSGS; among them are *children of Deaf adults* (CODAs), sign language interpreters, teachers, social workers, and persons otherwise interested in the language (Boyes Braem, 2012b).

DSGS has no standardized form but is composed of five dialects that originated in former schools for the Deaf, resulting in a Zurich, Berne, Basel, Lucerne, and St Gallen dialect. The differences

[1] *Grammar* here refers to an implicit set of rules that state how elements of the vocabulary are combined, not to a reference grammar. As Palfreyman, Sagara, & Zeshan (2015) write, "to date, not a single reference grammar of a sign language has been published that meets the common standards set by spoken language reference grammars" (p. 179). More recently, such a reference grammar has been released for New Zealand Sign Language (McKee, 2015).

between the dialects are primarily lexical and pertain to semantic fields such as food (distinct signs for regional food items) or date specifications (distinct signs for weekdays and months) (Boyes Braem, 1983). Figure 2.1 shows the example of the sign BROT ('BREAD') in the five dialects.[2]

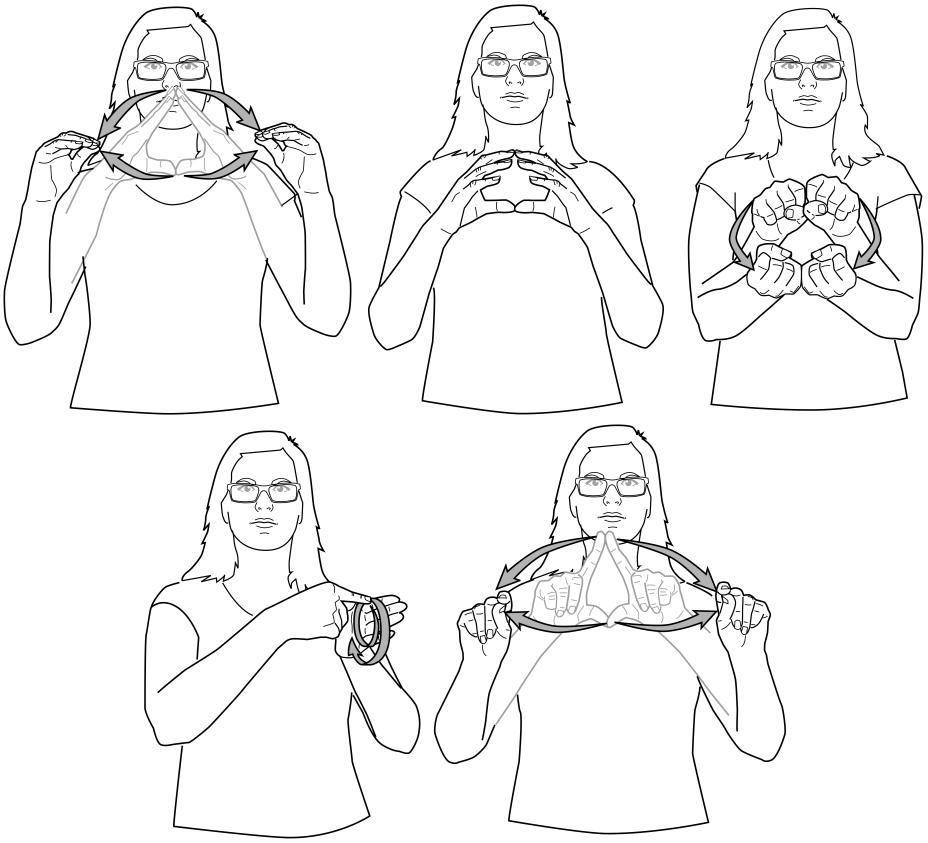


Figure 2.1: Sign BROT ('BREAD') in the five DSGS dialects (from top to bottom, left to right) Berne, Basel, Lucerne, St Gallen, and Zurich (figures from Boyes Braem, 2014)

2.1 Articulators

In terms of the reception and production channels involved, sign languages are *visual-gestural languages*, while spoken languages are *aural-oral languages*. More precisely, utterances in sign languages are produced with the hands/arms (the *manual activity*) and the shoulders, head, and face (the *non-manual components*). Manual and non-manual components together are known as the *sublexical components* of signs.

[2] "BROT" is an example of a *sign language gloss*, a label for one aspect of the meaning of a sign. Glosses are typically written in all caps. They are discussed in more detail in Section 2.2.

Stokoe (1960) was the first to divide the manual activity of signs into the parameters hand shape (the form of the hands, e.g., a fist, flat hand, etc.), location (where the manual activity is performed), and movement (an optional motion inherent in the sign). Later, researchers added a fourth component, hand position (the orientation of the hand) (Battison, 1978; Klima & Bellugi, 1979). These four components, which are now a common way of describing the manual activities of signs, are comparable to phonemes in spoken languages in that they can produce distinctions in meaning.[3] For example, in DSGS, the two signs SAGEN ('SAY') and FRAGEN ('ASK') shown in Figure 2.2 constitute a minimal pair. They differ in their hand shape.

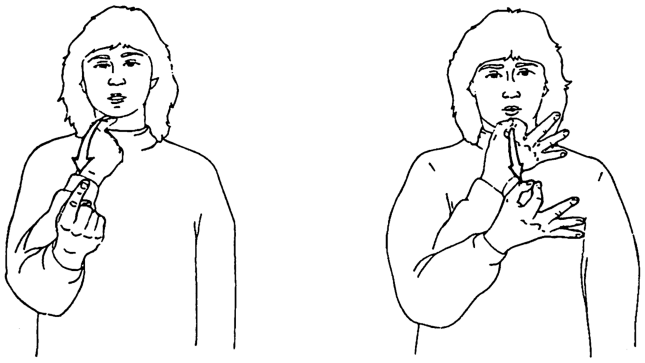


Figure 2.2: Minimal pair in DSGS: SAGEN ('SAY') (left) and FRAGEN ('ASK') (right) (figure from Boyes Braem, 1995)

The non-manual components of signing (such as head and shoulder movements, eyebrow movements, direction of eye gaze, etc.) are capable of assuming functions at all linguistic levels (Crasborn, 2006). It is therefore important to include non-manual information in automatic sign language processing, something which has not been done frequently in previous work. Chapters 4 and 5 deal with ways of incorporating non-manual information in sign language machine translation and sign language animation, respectively.

Pfau & Quer (2010) give examples of grammatical functions of non-manual articulations: determining sentence type, marking topicalized constituents, accompanying different types of embedded clauses, or expressing agreement and person distinctions in pronominals. As an example from DSGS, the sign sequence DU GEHÖRLOS ('YOU DEAF') can be taken to mean *Du bist gehörlos.* ('You are deaf.', declarative) or *Bist du gehörlos?* ('Are you deaf?', interrogative),

[3] It has been argued that the term *phoneme* is not appropriate for sign languages, since no phones (sounds) are involved in the production of signs. The term *chereme*, derived from the Greek word for 'hand', has been proposed as an alternative but has not reached wide acceptance.

depending on whether it is accompanied by a combination of non-manual components: To mark the sequence as interrogative, a signer slightly tilts the head forward, raises the eyebrows, and opens the eyes wide. This is shown in Figure 2.3.

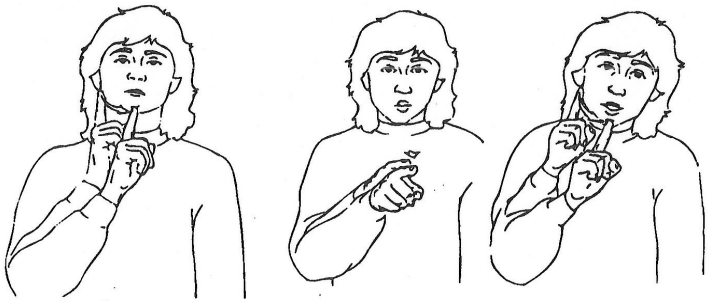


Figure 2.3: Non-manual information in DSGS: *Du bist gehörlos.* ('You are deaf.', left) vs. *Bist du gehörlos?* ('Are you deaf?', right) (figure from Boyes Braem, 1995)

As a further example, the German sentence *Der Bahnverkehr im Bahnhof Lenzburg ist beeinträchtigt.* ('Rail traffic in Lenzburg station is disrupted.') translates into DSGS as BAHNHOF LENZBURG IX BAHNVERKEHR BESCHRÄNKEN ('STATION LENZBURG IX RAIL-TRAFFIC DISRUPT'),[4] where BAHNHOF LENZBURG represents a topicalized constituent. Topicalization in DSGS is marked non-manually by raising the eyebrows and pushing the head forward (Boyes Braem, 1995). Conditional *if*/*when* utterances in DSGS have the head tilt and move forward slightly and the eyebrows go up at the start of the condition part. For rhetorical questions in DSGS, the head tilts and moves forward slightly and the eyebrows are furrowed on the question sign (Boyes Braem, 1995).

Research shows that the starting and ending times of non-manual components that serve linguistics functions, such as the ones described above, align with the boundaries of manual activities of signs. In Section 4.5, this observation will serve as a theoretical basis for viewing the task of automatically generating non-manual information as one of labelling glosses (as representations of manual components) with non-manual information. Non-manual components that serve purely affective purposes, e.g., expressing anger or disgust, are known to start slightly earlier than the surrounding manual components (Reilly & Anderson, 2002; Wilbur, 2000).

A special case of non-manual components are *mouthings* and *mouth gestures*. Mouthings are related to words of the most closely corresponding spoken language, i.e., German words for

[4] IX denotes an *indexical* (pointing) sign. There are different types of indexical signs. The qualifying attribute for this instance has been removed to increase readability.

DSGS, German Sign Language (*Deutsche Gebärdensprache*) (DGS), and Austrian Sign Language (*Österreichische Gebärdensprache*) (ÖGS), English words for American Sign Language (ASL), British Sign Language (BSL), and Irish Sign Language (ISL), etc. For example, the sign LAUTSPRECHER ('LOUDSPEAKER') in DSGS is accompanied by the German mouthing /Lautsprecher/. The spoken language words that serve as the basis for the mouthings are sometimes reduced to the part of the pronunciation that is visible on the lips or, in case of multisyllabic words, to the first few syllables (Boyes Braem, 2001b). Not every manual activity is accompanied by a mouthing; mouthings most frequently occur with nouns and verbs. When present, they can serve different functions, such as

- distinguishing between two signs with identical manual activity, e.g., BRUDER ('BROTHER') and SCHWESTER ('SISTER') in DSGS;
- restricting the meaning of a sign, e.g., the mouthing /Hamburger/ accompanying the sign FLEISCH ('MEAT') to express the concept *Hamburger* ('hamburger') in DSGS (Boyes Braem, 1995); or
- adding emphasis.

DSGS makes heavy use of mouthing: According to Boyes Braem (2001a), 80-90% of signs in DSGS are accompanied by a mouthing. Consequently, mouthings are highly frequent in the DSGS train announcements dealt with in this thesis.

Mouth gestures are mouth movements that are not related to spoken language words. They are produced with teeth, jaw, lips, cheeks, or tongue. Mouth gestures most commonly serve adjectival or adverbial function. For example, puffed cheeks in DSGS indicate that something comes in a large quantity.

The fact that sign languages are capable of expressing meaning through manual and non-manual components at the same time has led researchers to refer to them as *simultaneous* (or, *parallel*) languages in contrast to the *sequential* spoken languages. Simultaneity in sign languages refers not only to the co-occurrence of manual and non-manual components, but also to the possibility of using multiple non-manual components at the same time. An example of this was shown in Figure 2.3, where the eyebrows and the head worked together to mark a sign sequence as interrogative.

2.2 Notation systems

As was shown in Chapter 1, automatic sign language processing involves dealing with a written representation of signs. To date, no single widely accepted writing system for sign languages exists. A common way of providing a written record of signs, used throughout this thesis, are *glosses*. Glosses provide semantic labels of signs. They typically take on the base form of a word in the most closely corresponding spoken language. As with mouthings, the spoken language used for DSGS, DGS, and ÖGS glosses is German. Glosses are typically written in all caps. As an example, the gloss GESCHWISTER ('SIBLINGS') is used to represent the DSGS sign for the concept *Geschwister* ('siblings').

Glosses provide an easily readable and searchable representation of signs, allowing, in particular, for alphabetic sorting in a lexicon (Boyes Braem, 2012a). However, from a conceptual point of view, expressing the vocabulary of one language (a sign language) by means of another (a spoken language) is problematic as it involves a translation step (Pizzuto, Rossini, & Russo, 2006). A further problem with glosses is that they are usually not standardized for a particular sign language: In principle, the same sign may be denoted with different glosses, and the same gloss may be used to denote different signs. This issue can be circumvented by assigning glosses in a controlled manner (i.e., introducing a new gloss only if an appropriate one is not available in a set of glosses). This was done for the work reported in this thesis.[5]

An important remaining issue is that glosses convey only limited non-manual information. With the exception of the use of mouthings and mouth gestures, few signs have mandatory non-manual components at the lexical level. It is mostly at other linguistic levels that non-manual information comes into play. Hence, representing a signed utterance merely with glosses in most cases involves encoding information about the manual activity of a sign only. It is clear that such a representation falls short of capturing signing as what it is: a multi-level phenomenon composed of manual and non-manual information. Ideally, therefore, glosses should be complemented with non-manual information on separate tiers. An example of such a representation is shown in Table 2.1: The gloss tier is complemented with two tiers providing information about movements

[5]Note that this is not equivalent to the use of *ID glosses* (Johnston, 2008). ID glosses also subsume phonological and morphological variants of a lexeme under the same gloss. This was not a desired property for the work reported in this thesis, since morphological and phonological variants have different forms and one gloss entry was needed for each distinct form in anticipation of the sign language animation step.

of the eyebrows and the head, respectively.

Gloss	MELDUNG (‘NOTICE’)	IX (‘IX’)	BAHN (‘TRAIN’)	S1 (‘S1’)	NACH (‘TO’)	LUZERN (‘LUCERNE’)	AUSFALL (‘CANCELLATION’)
Eyebrows	raised				neutral	raised	
Head	forward	back	up	down	up		down

Table 2.1: DSGS translation of the German train announcement *Ausfallmeldung zur S1 nach Luzern* (‘Notice of cancellation regarding the S1 to Lucerne’)

A notation system that takes account of the visual nature of sign languages is one that describes the physical form of signs. The Hamburg Notation System for Sign Languages (HamNoSys) (Prillwitz, Leven, Zienert, Hanke, & Henning, 1989) represents such a notation. For my work in automatic sign language processing as reported in the following chapters, I applied both glosses as a meaning-based notation system and HamNoSys as a form-based notation system.

HamNoSys is based on the categorization of manual components by Stokoe (1960) (cf. Section 2.1). The system consists of approximately 200 symbols describing the components hand shape, hand position (with finger direction and palm orientation as sub-components), location, and movement. The symbols together constitute a Unicode font. The current version is HamNoSys 4.0, with plans for version 5.0 being underway. Figure 2.4 shows the HamNoSys notation of the DSGS sign VOLK (‘PEOPLE’) that contains one instance of each manual component.

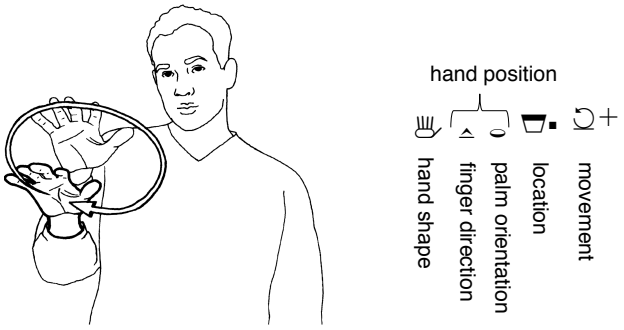


Figure 2.4: HamNoSys notation of the DSGS sign VOLK (‘PEOPLE’) (Boyes Braem, 2001c)

A drawback of HamNoSys notations is that they can grow long and complex, as becomes obvious from the notation of the DSGS sign GEBÄRDENSPRACHKURS (‘SIGN LANGUAGE COURSE’) in Figure 2.5. It is therefore easy to see why HamNoSys is not used as a daily writing system by Deaf signers.

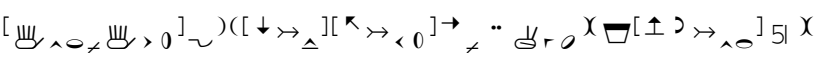


Figure 2.5: HamNoSys notation of the DSGS sign GEBÄRDENSPRACHKURS (‘SIGN LANGUAGE COURSE’) (Boyes Braem, 2001c)

SignWriting (V. Sutton, 2010) lends itself more to this purpose, though it is not widely applied on a day-to-day basis, either, the important point being that signers do not frequently produce written records of signing in their everyday life. SignWriting consists of approximately 640 symbols grouped into the seven categories hands, movement, face/head, body, dynamics and timing, punctuation, and advanced sorting. Figure 2.6 shows the SignWriting notation of the DSGS sign GEBÄRDENSPRACHKURS ('SIGN LANGUAGE COURSE').[6] SignWriting symbols are arranged in a "two-dimensional space as a map of a human body" (van der Hulst & Channon, 2010, p. 159). Thus, in this notation system, location is not expressed explicitly through dedicated symbols (as in HamNoSys) but by way of placement in the 2D space. This poses challenges for automatic processing of SignWriting. Some of these challenges have been resolved in a new version of the notation system, Modern SignWriting.[7] However, it is still safe to say that of the two form-based sign language notation systems described in this section, HamNoSys is the one that is better equipped for automatic processing.

Figure 2.6: SignWriting notation of the DSGS sign GEBÄRDENSPRACHKURS ('SIGN LANGUAGE COURSE')

2.3 Word order

As noted at the beginning of this chapter, sign languages have their own grammars, which are not identical to the grammars of the surrounding spoken languages.[8] This is an important observation for machine translation between sign languages and spoken languages, as it means that the possibility of constituent reordering always needs to be present in such a translation system.

At the beginning of this chapter, it was pointed out that a reference grammar "that meets the common standards set by spoken language reference grammars" (Palfreyman et al., 2015, p. 179)

[6] `http://www.signbank.org/signpuddle2.0/searchword.php?ui=8&sgn=48&sid=1226&sTrm=*&type=&sTxt=&sSrc=&` (last accessed October 1, 2015).

[7] `https://github.com/Slevinski/msw` (last accessed December 16, 2015).

[8] Communication systems that adhere to the grammars of the surrounding spoken languages are discussed in Section 2.6.

exists only for New Zealand Sign Language (McKee, 2015). Nevertheless, constituent order as one part of the grammar has been researched for some sign languages. All of these sign languages have been shown to exhibit either subject–object–verb (SOV) or subject–verb–object (SVO) order (Leeson & Saeed, 2012). For example, DGS and Sign Language of the Netherlands (*Nederlandse Gebarentaal*) (NGT) have been shown to follow SOV order (Coerts, 1994; Erlenkamp, 2001). Conversely, ASL and Hong Kong Sign Language exhibit basic SVO order (Fischer, 1975; Sze, 2003).

DSGS has been shown to license both SVO and SOV order (Boyes Braem, 2005). For example, both of the following DSGS translations of the German sentence *Der Hund holt den Knochen.* ('The dog fetches the bone.') are correct: HUND HOLEN KNOCHEN ('DOG FETCH BONE') and HUND KNOCHEN HOLEN ('DOG BONE FETCH'). SOV order is typically applied in cases where there is no risk of confusing the object with the subject as well as in combination with directional signs, e.g., GEHEN ('GO') in ICH KINO GEHEN ('I CINEMA GO').

2.4 Iconicity

In spoken languages, the relation between form and meaning is arbitrary for most words; examples of exceptions are *onomatopoeia*. By contrast, sign languages feature many signs that are iconic. As Johnston & Schembri (2007) note, "this greater degree of iconicity in visual-gestural languages is not particularly surprising because objects and actions in the external world tend to have more visual than auditory associations" (p. 3). Iconicity is taken into account for a lexical comparison of DGS and DSGS as reported in Section 2.4.3.

Although iconicity is a unifying principle across sign languages, sign languages differ more strongly in their lexicon than in their grammar (Boyes Braem, 1995). This is because iconicity can draw on different aspects of the form-meaning relation. In other words, even if two meaning-equivalent signs from two different sign languages are iconic, their underlying images are not necessarily the same. For example, the signs for 'TREE' in Taiwanese Sign Language (TSL) and Chinese Sign Language (CSL) are both iconic but have different underlying images: The TSL sign depicts a tree as a whole, while the CSL sign visualizes only the trunk of a tree (Xu, 2006).

From Johnston (1989), based on work by Klima & Bellugi (1979), comes a distinction into four degrees of iconicity: Signs whose meaning can be deduced from the form via the underlying

image are considered to be *transparent*. An example is the DSGS sign HAUS ('HOUSE'): It is performed by sketching the outline of a house. Signs like the DSGS sign EINVERSTANDEN ('AGREED'), performed by moving the dominant hand from the left shoulder to the right hip using a flat hand shape, are *opaque* in that the relation between form and meaning is not obvious even if the meaning is known. Between transparency and opaqueness lie *translucency* (the relation between the form and the meaning of a sign becomes clear once the meaning is known) and *obscureness* (the sign has an underlying image, but the relation between form and meaning is not clear).

A frequently observed tendency in sign languages is for initially iconic signs to lose part of their iconic value (Sandler & Lillo-Martin, 2006). The iconic value of signs may fade, for example, through a change in signing location, a change from non-symmetry to symmetry, or the reduction of two segments of a sign to one (Frishberg, 1979).

2.4.1 A typology of signs based on iconicity

A typology of signs that takes iconicity into account distinguishes between *conventional*, *productive*, and other signs (Johnston & Schembri, 1999) as shown in Figure 2.7. Conventional signs are idiomatic in that their overall meaning is not composed solely of the meanings of their sublexical components. These are signs found in a lexicon, which renders them similar to spoken language words. Most conventional signs were originally iconic, yet through one of the processes mentioned above have developed into form-meaning units that can be used without drawing on the initial iconic value alone. An example of a conventional sign in DSGS is KAISERSCHNITT ('C-SECTION'): While the sum of the meanings of the sublexical components is that of a longish object (a knife) moving along the lower part of the body, the overall meaning of the sign is more specific in that it refers to a particular medical procedure that involves cutting with a knife, the c-section. The iconic value of many conventional signs can be reactivated by modifying the signs, e.g., pluralizing them. This process is called *delexicalisation*, or *re-iconisation*.

By contrast, productive signs are signs whose meanings are composed of the sum of the meanings of their sublexical components only. Productive signs are always iconic and are derived spontaneously. Productive signs do not have a stable citation form; hence, they do not appear in a sign language lexicon. However, they are abundant in everyday signing, especially in narratives. Productive signs can turn into conventional signs through the process of *lexicalisation*.

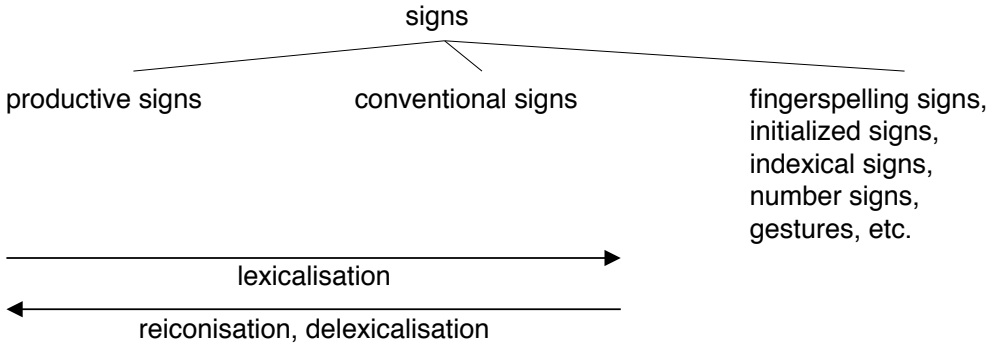


Figure 2.7: Typology of signs based on Johnston & Schembri (1999)

The third category of signs as shown in Figure 2.7 contains sub-categories such as fingerspelling signs, initialized signs, indexical signs, number signs, gestures, etc. Fingerspelling signs are introduced in Section 2.5. The remaining sub-categories lie outside of the scope of this thesis.

2.4.2 Image-producing techniques

A rare but possible case is for two iconic signs to share the same underlying image but employ a different "*image-producing technique*" (Konrad, 2013, p. 115). For example, DGS features two signs for KRIPPE ('CRIB') that have the same underlying image, that of a crib, but different forms: In the case of the first sign, the hands and arms symbolize the two legs of the crib (and through this, the crib as a whole), while in the case of the second sign, the image of the crib is evoked by tracing its outline (Konrad, 2011). When assessing whether two signs have identical iconic values, it is essential to look at not just a general description of the underlying image (such as "crib") but to precisely determine the image-producing techniques at play. This is the approach pursued for a lexical comparison of DGS and DSGS as described in Section 2.4.3.

Six image-producing techniques exist: the substitutive, manipulative, sketching, stamping, measuring, and indexing technique (Langer, 2005). These techniques describe the function of the hand shape and of the motion optionally inherent in a sign. In two-handed signs, each hand may employ a different technique. In the *substitutive technique*, the hand represents the whole or part of an object. The example of the hands/arms representing the legs of a crib for the DGS sign previously introduced is an instantiation of this technique. As a further example from DGS, in Figure 2.8, the hands represent the bars of a prison window to convey the image of a prison. In the *manipulative technique*, the hand represents itself, i.e., the hand of a person, and performs actions like handling or manipulating an object. In Figure 2.8, the hand is moving a fishing rod,

depicting the image of fishing. As part of the *sketching technique*, the hand traces the shape of an object, e.g., a crib as in the DGS example previously mentioned. In Figure 2.8, the hand outlines the shape of a tube to evoke the image of precisely that object. The *stamping technique* is characterized by the hand imprinting a pattern on a surface. For example, in Figure 2.8, the hand stamps the lines of an imagined regulatory document to induce the image of a rule. The *measuring technique* involves specifying the dimensions of an object. For example, in Figure 2.8, the hand indicates a person's small height to express the image of a child. Finally, the *indexing technique* consists of pointing at an object to generate the image of that object, such as the image of a nose in the example in Figure 2.8.

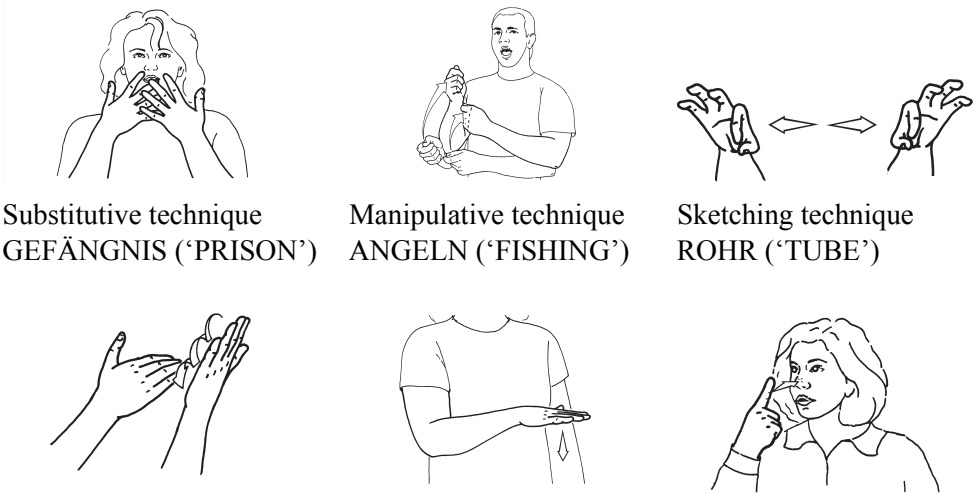


Figure 2.8: Image-producing techniques in DGS (figures from Langer, 2005)

2.4.3 A lexical comparison of DGS and DSGS

[This section is an extension of Ebling, Konrad, Boyes Braem, & Langer (2015).]

DGS and DSGS are used in geographically adjoining regions, and most of their users have German as a second language. German is also used as a source for mouthings (cf. Section 2.1) and fingerspelling (cf. Section 2.5) in both sign languages. In an ongoing study, we are investigating the lexical similarity between DGS and DSGS. While the similarity between the two languages is anecdotally reported as being high (Boyes Braem, Haug, & Shores, 2012), ours is the first study to establish lexical overlap empirically. Being a linguistic project in its own right, the study also offers the potential of drawing on insights gained and resources developed for DGS. As will be shown in Chapter 3, presumably the largest sign language corpus currently being built is the

DGS Corpus, a long-term project of the Academy of Sciences in Hamburg, Germany. At the end of the 15-year project in 2023, the corpus is expected to hold 3.5 million tokens originating in 6 400 hours of video of signing from 330 informants (Hanke, 2013). It will serve as the basis for an electronic DGS/German dictionary. DSGS is comparatively resource-poor. The possibility of leveraging linguistic analyses and notations for lexical items from the DGS dictionary would have the potential of speeding up the process of building corpora for DSGS.

Previous studies undertaking lexical comparisons of other sign languages[9] looked at only the form parameters hand shape, hand position, location, and movement (cf. Section 2.1) to establish lexical similarity. In contrast, our approach takes into account both form and iconicity. For iconicity, our approach relies on an analysis of the image-producing technique (cf. Section 2.4.2) for each hand involved in the production of a sign. In doing so, it goes beyond previous work that took into account only the general notion of iconic motivation (S. Su & Tai, 2009; Xu, 2006). The more precise concept of image-producing techniques is advantageous in cases where the same image can be constituted by different techniques, as in the example of the sign KRIPPE in DGS given in Section 2.4.2.

Existing lexical databases for DGS and DSGS serve as data for our analysis. In these databases, the meaning of each sign is described with several keywords. For example, associated with the sign FILM ('MOVIE') in both DGS and DSGS are the keywords "Film" and "Kinofilm". We automatically identified all pairs of DGS/DSGS signs that had at least one keyword in common, such as the DGS and DSGS signs FILM. From this set, we removed all geographical signs (e.g., the sign BERLIN), as many of these signs have been borrowed from other sign languages, which would result in an inflated percentage of signs from the two languages being the same. In addition, signs for body parts and pronouns were discarded, as these are largely indexical (pointing) signs in the two languages considered (such as the DGS sign NASE shown in Figure 2.8). We also eliminated number signs and fingerspelling signs. This was followed by a manual check of the remaining sign pairs to ensure that their meanings were indeed identical. The resulting set consisted of 648 concepts expressed in 1 818 pairs of DGS/DSGS signs.

Our comparison approach is visualized in Figure 2.9: Two meaning-equivalent iconic signs from the two languages are considered *lexically identical* if they have the same image-producing technique and form (path 1 in Figure 2.9). If two signs are not iconic, it is sufficient for them to have

[9] See Ebling, Konrad, et al. (2015) for an overview.

the same form in order to be considered lexically identical (path 5).

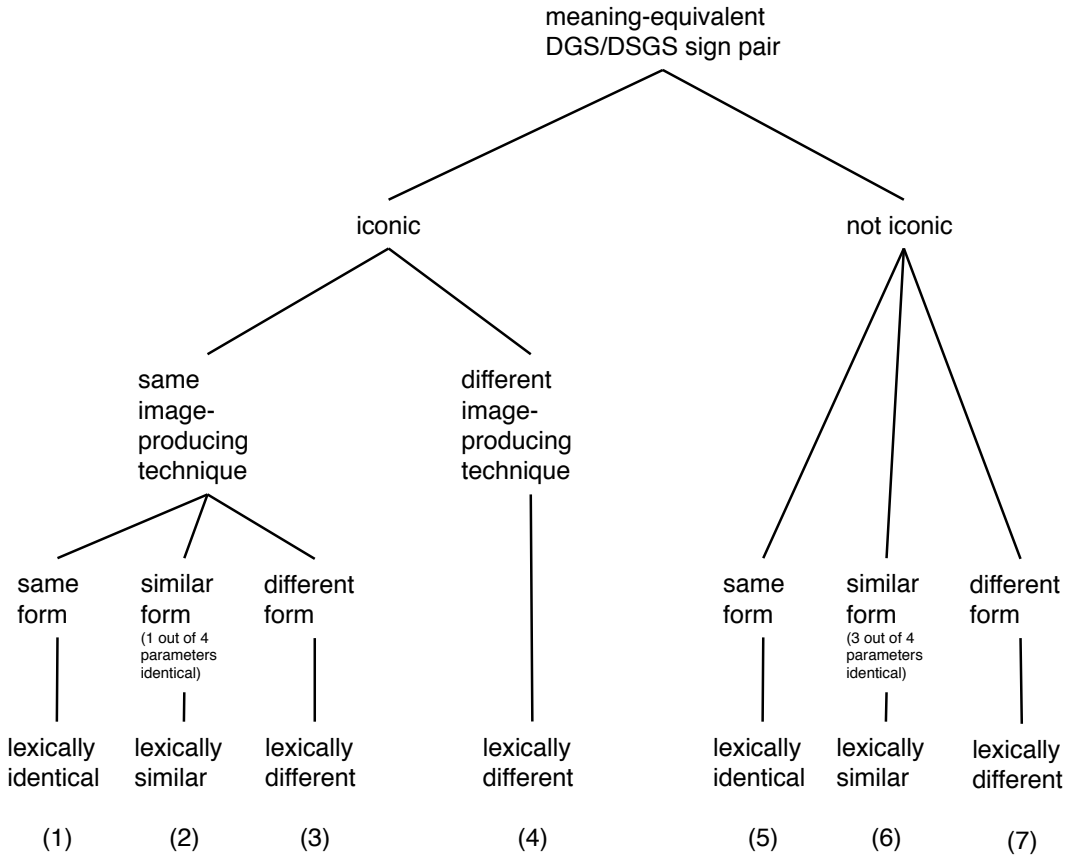


Figure 2.9: DGS/DSGS sign pair comparison

For two iconic signs to be *lexically similar*, they have to have the same image-producing technique and a similar form (path 2). Taking into account iconicity here allows us to be less strict with regard to form: If we know that the image-producing techniques of two signs are identical, it is sufficient to demand for one out of four form parameters (hand shape, hand position, location, movement) to be the same for the overall forms to be similar. If two signs are not iconic, they have to have all but one (i.e., three out of four) form parameter in common to be considered lexically similar (path 6).[10]

All other pairs are considered to be *lexically different*. This means that in particular, iconic sign pairs that have different image-producing techniques are rated as lexically different, irrespective of their form (path 4). Two signs that are iconic and have the same image-producing technique

[10]This is the criterion applied throughout in previous research, which took into account only form parameters.

but a different form are also rated as lexically different (path 3). The same applies for two non-iconic signs that have different forms (path 7).

While it is straightforward to automatically compare for image-producing techniques based on information from lexical databases, automatically comparing for form equivalence is not. Since we have HamNoSys notations (cf. Section 2.2) available for both the DGS and DSGS entries in our set of signs to be compared, we first attempted an automatic form comparison. However, we found that this was inherently difficult: HamNoSys strings can be written in more or less explicit ways and, due to the lack of orthography, also in different ways. We therefore automated only the search for identical notations and are comparing the remaining forms manually by looking at video recordings of the signs. This process is currently underway.

2.5 Fingerspelling

Apart from conventional and productive signs, a third category of signs exists that comprises various types of signs, as shown in Figure 2.7. Among them are *fingerspelling signs*. Sign languages make use of a communication form known as the *finger alphabet* (or, *manual alphabet*), in which the letters of a spoken language word are fingerspelled, i.e., dedicated signs are used for each letter of the word. The letters of the alphabet of the most closely corresponding spoken language are used, e.g., English for ASL, BSL, and ISL, German for DGS, ÖGS, and DSGS, etc. Figure 2.10 shows the manual alphabet of DSGS. Note that it features dedicated signs for -Ä-, -Ö-, -Ü-, -CH-, and -SCH-. Section 5.4 reports on work in synthesizing the finger alphabet of DSGS and evaluating the comprehensibility of the resulting fingerspelling sequences.

Some fingerspelling signs are iconic, i.e., their meaning becomes obvious from their form. Such is the case, e.g., with -C-, -L-, or -O- in DSGS. Most manual alphabets, like the one for DSGS, are one-handed, an exception being the two-handed alphabet for BSL.

Fingerspelling is often used to express concepts for which no lexical sign exists in a sign language, e.g., for proper names. As such, it is frequent in train announcements, where many place names occur.

Frequency of use and speed of fingerspelling vary strongly across sign languages. For example, ASL is known to make heavy use of the finger alphabet. In contrast, fingerspelling is less common in DSGS: Until recently, DSGS signers used mouthings to express technical terms or

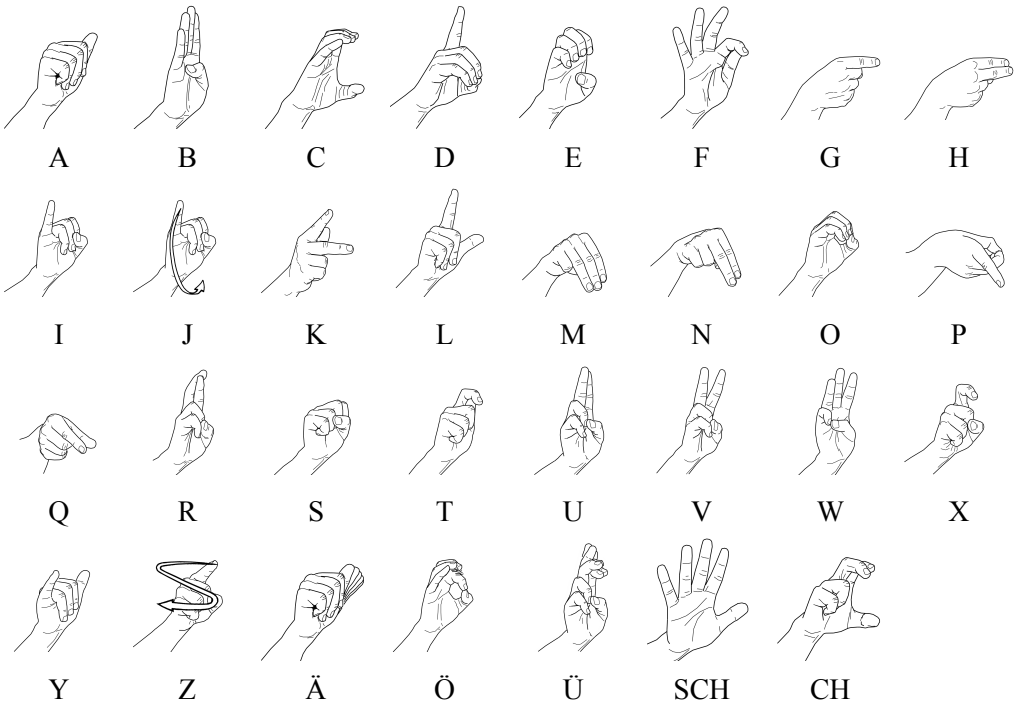


Figure 2.10: Finger alphabet of DSGS (Boyes Braem, 2001c)

proper names for which no lexical sign existed. Nowadays, fingerspelling is used more often in these cases, particularly by younger DSGS signers. In addition, fingerspelling is applied with abbreviations: For example, the abbreviation I-C for *InterCity*, a type of train, is fingerspelled.

Keane & Brentari (2015) report fingerspelling rates between 2.18 and 6.5 letters per second (with a mean of 5.36 letters per second) based on data from different sign languages. The speed of ASL fingerspelling is known to be particularly high (Padden & Gunsauls, 2003), whereas fingerspelling in DSGS is much slower. In our study on the comprehensibility of synthesized fingerspelling sequences (cf. Section 5.4), a human signer performing the same fingerspelling sequences as a signing avatar exhibited a fingerspelling rate of 1.76 letters per second.

2.6 Communication systems similar to sign languages

In contrast to sign languages developed in Deaf communities, there are other forms of communication that involve signs. For example, International Sign (IS) is a communication system used at international events where interpreters for some sign languages might not be available. There is no standardized form of IS; the system varies depending on the preferred sign language of the individual interpreter but in all cases utilizes many iconic techniques.

A communication system similar to DSGS is *Signed German based on DSGS*:[11] In this system, DSGS sign correspondences for each word of a German sentence are signed in German word order. Accordingly, the system is called *lautsprachbegleitendes Gebärden* ('spoken-language-accompanying signing') in German. A similar system exists for English as *Signed English* in combination with ASL, BSL, and ISL as well as for other spoken language/sign language pairs. Within these systems, signs that do not exist in the corresponding sign languages are introduced, such as signs for spoken language determiners and some conjunctions. An example of a sentence in Signed German based on DSGS is shown in Figure 2.11 along with the corresponding DSGS sentence. Note that the DSGS sentence follows SOV order as mentioned in Section 2.3. The corresponding German sentence is *Ich habe eine Milchflasche aus dem Kühlschrank genommen.* ('I have taken a bottle of milk out of the refrigerator.'). In the example, the Signed German sentence uses dedicated signs for the spoken language word forms *habe* ('have'), *eine* ('a'), *aus* ('out of'), and *dem* ('the') as well as the morpheme *ge-* (past tense marker), none of which are present in the DSGS sentence.

For my experiments in automatically translating from German to DSGS (cf. Section 4.4), I used a baseline similar to Signed German based on DSGS.

2.7 Summary

This chapter has laid the linguistic foundations for the following chapters, which deal with the automatic processing of sign languages. I have introduced the articulators in signing and have shown, in particular, that non-manual components make up an important part of signing in that they assume linguistic functions. This observation underlines the need for including non-manual information in an automatic sign language processing pipeline involving machine translation and sign language animation.

The chapter has also given an overview of different sign language notation systems. My work reported in this thesis uses both a meaning-based and a form-based notation system. More precisely, glosses are used as sign language representations in the machine translation task, while information about the physical form of the signs (through HamNoSys notations) is employed in the subsequent animation task.

[11] Signed German also exists on the basis of DGS and ÖGS.

Knowledge of word order in sign languages, as conveyed in this chapter, prepares the ground for an understanding of the prerequisites of automatic translation between spoken languages and sign languages. As has been shown, sign languages do not follow the word order of the surrounding spoken languages. However, there are communication systems (as opposed to fully natural languages) for which precisely that is true.

The concept of iconicity has been introduced. I have shown that the spectrum from an arbitrary to an iconic form-meaning relation is typically divided into four gradations: opacity, obscureness, translucency, and transparency. Iconicity is also the basis for distinguishing between conventional and productive signs. Iconicity is more prevalent in sign languages than in spoken languages. It is therefore essential to take this concept into account when undertaking a lexical comparison of sign languages, something which has recently been initiated for DGS and DSGS. Since there is a possibility of two signs having identical underlying images but different forms, it is necessary to take into account not just general descriptions of underlying images, but to look more closely at the techniques through which iconicity is constituted. Six image-producing techniques are commonly assumed. Our lexical comparison of DGS and DSGS relies on an analysis of these techniques, thereby extending previous work that considered only the more general concept of iconic motivation.

This chapter has also introduced fingerspelling. I have shown that fingerspelling is used most often in situations where no lexical sign exists for a concept. In particular, it is often applied to place names. As such, fingerspelling is common in signed train announcements. Later in this thesis, I will report on work in synthesizing the finger alphabet of DSGS and evaluating the comprehensibility of the resulting fingerspelling sequences.

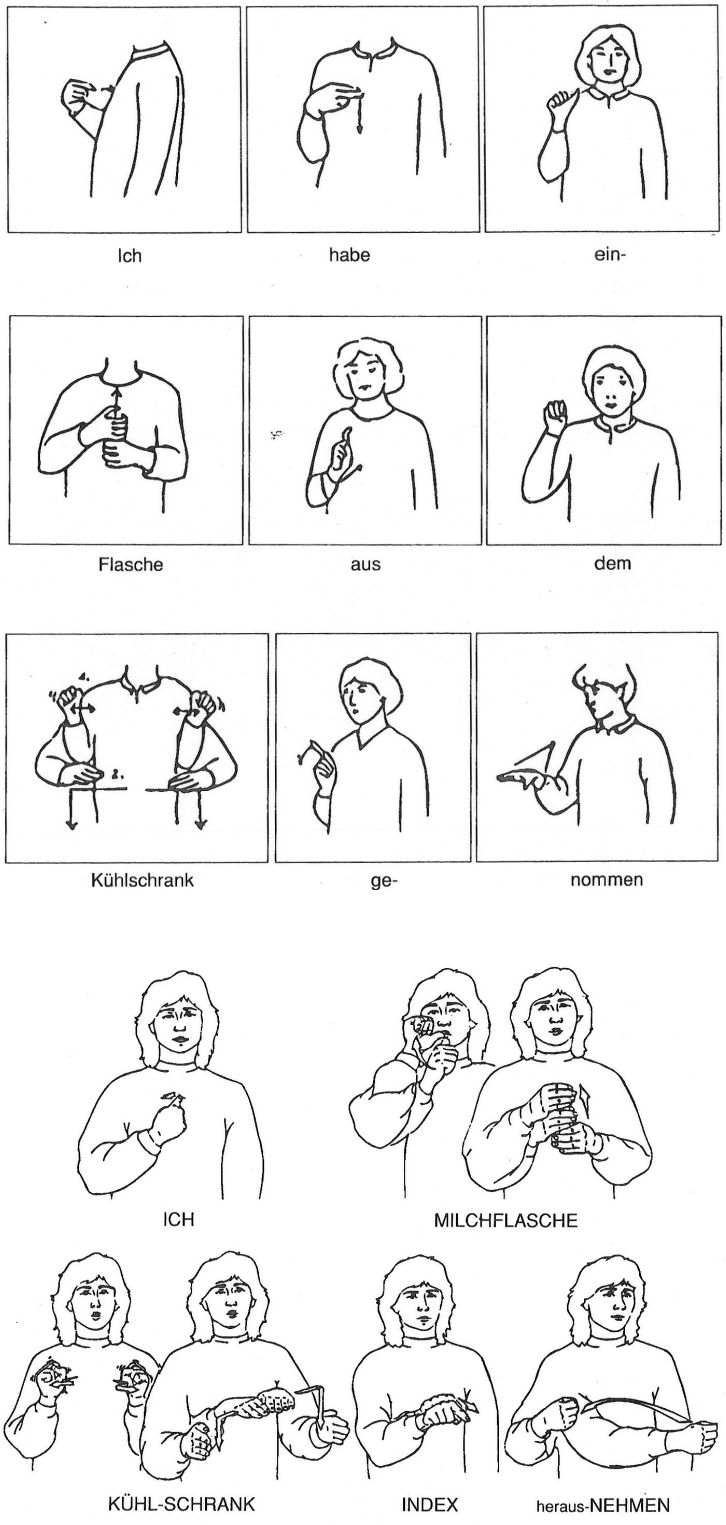


Figure 2.11: Sentence in Signed German based on DSGS (top) and DSGS (bottom) (figures from Boyes Braem, 1995)

Chapter 3

Sign language corpora

[This chapter is an extension of Ebling (2013) and Ebling (2016).]

Owing to the availability of increased computing power and the possibility of processing ever larger amounts of data, corpus linguistics has received growing attention since the mid-1980s. A *corpus* is a collection of language data available in electronic form. Optionally, a corpus may also contain metadata and linguistic annotations (Lemnitzer & Zinsmeister, 2006). Several corpus definitions additionally emphasize the notion of representativeness, i.e., the property that "findings based on [a corpus'] contents can be generalized to a larger hypothetical corpus" (Leech, 1991, p. 27). For example, McEnery & Wilson (2001) define a corpus as a "finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration" (p. 32). A corpus according to Tognini-Bonelli (2001) is "a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis" (p. 2). Sinclair (2005) describes a corpus as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (p. 16).

Sign languages are low-resource languages, which means they often lack the resources available for many spoken languages, such as, in particular, corpora. For some sign languages, corpora do exist.[1] Presumably the largest sign language corpus currently being built is the German Sign Language (*Deutsche Gebärdensprache*) (DGS) Corpus mentioned in Section 2.4.3, which is

[1] See `http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/gs-korpora.html` for a comprehensive overview (last accessed October 1, 2015).

estimated to consist of 3.5 million tokens upon completion. Sign language corpora are *multimodal corpora* (Allwood, 2009) in that they consist of data from more than one modality, where modality refers to either a sensory or a production modality. With sign languages, more than one production modality is always at play due to the fact that information is conveyed via both manual and non-manual articulators (cf. Section 2.1).

In what follows, the process of building a sign language corpus is described. At the very least, it consists of obtaining *raw data* (Section 3.1) and creating *primary data* (Section 3.2). Optionally, *secondary data* (Section 3.3) and *metadata* (Section 3.4) may be added (Konrad, 2011).

3.1 Obtaining raw data

Obtaining raw data for a sign language corpus consists of collecting existing or producing new video recordings of signing. When recording sign language, multiple video cameras are sometimes used to allow for bird's eye, profile, and close-up views, e.g., of the facial expressions of signers. During acquisition of data for the DGS Corpus (Nishio et al., 2010), up to eleven cameras (including 3D cameras) were used to record two informants at a time (Hanke, 2013).

If at the time of recording informants are aware of the fact that they are contributing data for research on language usage, the *observer's paradox* described by Labov (1972) comes into play: "The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation." (p. 209) Hence, the informants' *linguistic self-awareness*, i.e., the extent to which they pay attention to their language usage, is an important aspect in the process of obtaining raw data for any corpus. Himmelmann (1998) identified four types of communicative events that each correspond to a degree of linguistic self-awareness:

- *Natural communicative events*: These are events in which the informants have no linguistic self-awareness. Such events are nearly undocumentable, as the mere presence of a recording device or observer impedes full naturalness.
- *Observed communicative events*: These events exhibit the most natural use of language given the presence of a recording device or observer.
- *Staged communicative events*: These events are put on for the sole purpose of generating language data.

- *Elicitation*: These events are the least natural. They are highly structured: Informants are asked to embed a linguistic unit (e.g., a sign) in context, translate it into their native language, or provide acceptability judgments.

Many sign language corpora created for use in linguistic research contain data from staged communicative events;[2] examples are the previously mentioned DGS Corpus (Nishio et al., 2010), the Signs of Ireland Corpus (Leeson, Saeed, Macduff, Byrne-Dunne, & Leonard, 2006), or the Corpus NGT (Crasborn & Zwitserlood, 2008). During staged communicative events, informants are asked to, e.g., share their opinion on a controversial topic, coordinate calendars, or establish a narrative based on pictures or videos. The choice of tasks and topics governs the output: For example, having informants debate about topics related to politics or society is likely to yield many first-person statements, while asking them to discuss the possible meanings of traffic signs results in more objective language. Also common are *map tasks*: Here, two informants sit opposite each other, each with a map in front of them. One informant assumes the role of the instruction giver, the other the role of the instruction follower. Both maps have landmarks on them, the instruction giver's map additionally shows a route to be followed. The landmarks may differ between the two maps to increase the likelihood of specific lexical items being produced.

3.2 Creating primary data

Following the collection of existing or the production of new video recordings, the continuous signing stream in these recordings is *segmented* and *notated* to create the primary data. These two steps are commonly comprised under the term *transcription*. A translation into spoken language (usually sentence by sentence) may be provided along with the transcription.

3.2.1 Segmentation

Segmenting a signed utterance consists of splitting it into individual units, which are typically signs. This renders segmentation in sign language corpora similar to tokenization in spoken language corpora. Two different approaches to determining the beginning and the end of a sign

[2] See Hong et al. (2009) for an overview.

exist: broad and narrow segmentation (Hanke, Matthes, Regen, & Worseck, 2012). Segmenting broadly means regarding the end of the previous sign as the beginning of the current sign, i.e., considering the transition from the previous to the current sign as part of the current sign. Conversely, in narrow segmentation, the beginning of the current sign is not assumed until all of the manual components except for the movement (i.e., hand shape, hand position, and location; cf. Section 2.1) are in place. This approach treats the transitions between signs as segments of their own. Figure 3.1 shows an example of determining the beginning and the end of the sign BIS ('TO') in the Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS) translation of the German date specification *5. bis 7. September* ('September 5 to 7'), FÜNFTE BIS SIEBTE SEPTEMBER ('FIFTH TO SEVENTH SEPTEMBER'). Segmenting in a wide manner implies determining the beginning of the sign BIS to correspond to the end of the sign FÜNFTE and the end of the sign BIS to correspond to the point in time when the movement of the sign has been completed. This is shown in the upper part of Figure 3.1. By contrast, in narrow segmentation, the sign BIS does not start until all parameters except for the movement are in place. The end of the sign is the same as in wide segmentation. This is displayed in the lower part of Figure 3.1.

Figure 3.1: Wide (above) and narrow (below) segmentation of the DSGS sequence FÜNFTE BIS SIEBTE SEPTEMBER ('FIFTH TO SEVENTH SEPTEMBER')

The decision as to whether broad or narrow segmentation is applied is largely determined by the research question underlying the creation of a sign language corpus. This decision has implications for the subsequent notation step: Typically, only the segments identified as signs are

considered for notation, and any transitional segments are ignored. As will be shown in Section 3.6, we applied narrow segmentation in our corpus of DSGS train announcements.

3.2.2 Notation

Following segmentation, the segments identified as signs are usually labelled, i.e., a written record of the signs is provided. It was pointed out above that this step together with the previous segmentation step is commonly referred to as *transcription*. Sometimes, transcription is conceived more narrowly as referring to the step under consideration only. Here, this step by itself is referred to as *notation*.

In Section 2.2, reference was made to the fact that no standardized writing system for sign languages exists. Sign language glosses have been used as a way of labelling the meanings of signs. Mention was made of the fact that glosses most often encode information about the manual activity only. Non-manual information is typically added on separate tiers, as shown in the screenshots of a Greek Sign Language corpus in Figure 3.2, where a separate tier exists, e.g., for eyebrow and mouth gesture information.

Systems like the Hamburg Notation System for Sign Languages (HamNoSys) or SignWriting (cf. Section 2.2) record the physical form of signs. Depending on the purpose of a sign language corpus, sometimes both meaning- and form-based notation is carried out. Such was the case for our corpus of DSGS train announcements introduced in Section 3.6.

3.3 Creating secondary data

Creating secondary data for a sign language corpus consists of adding linguistic annotations. Recall that the presence of annotations is not a constitutive but rather an optional feature of a corpus. The range of linguistic phenomena that have been annotated in sign language corpora is wide and guided by the research question that motivated the creation of the sign language corpus. A good overview can be found in Konrad (2011). For the DSGS corpus of train announcements introduced in Section 3.6, no explicit linguistic annotations were added.

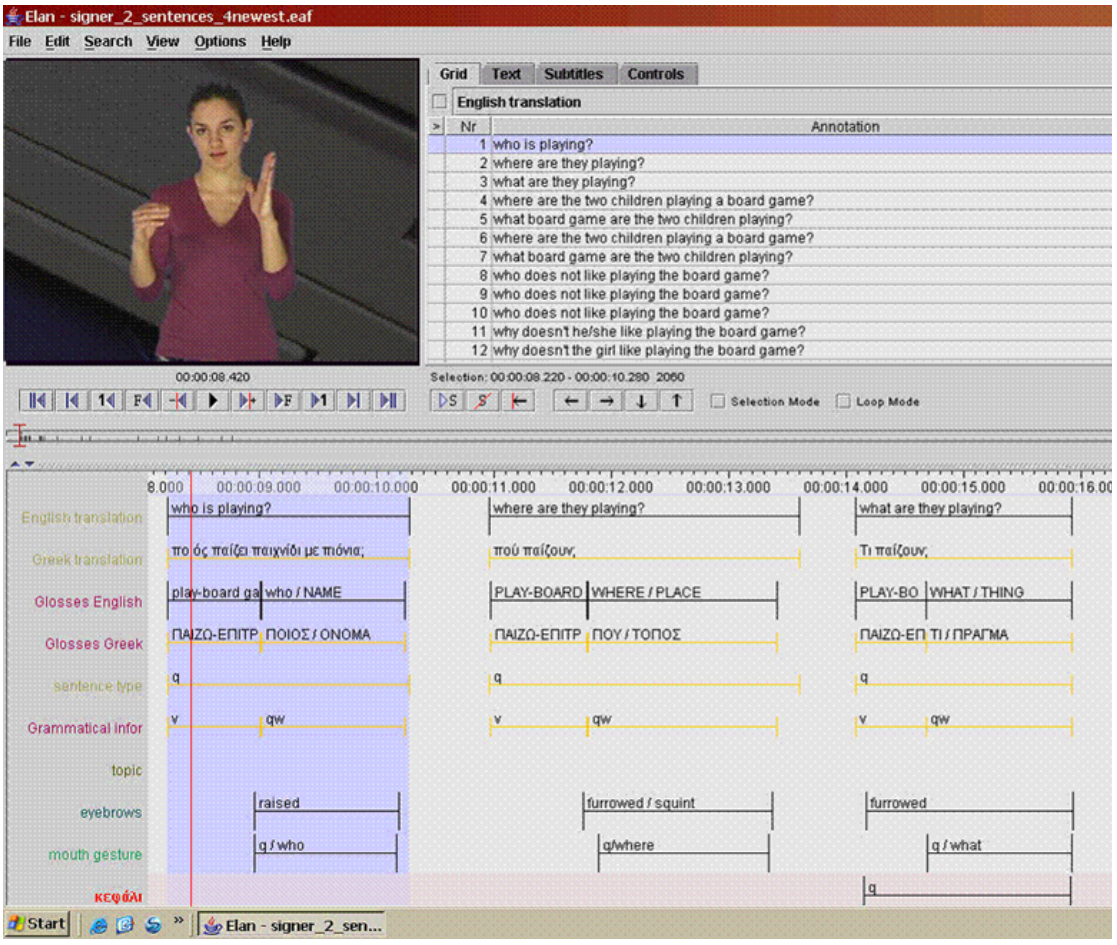


Figure 3.2: Greek Sign Language corpus in ELAN (figure from Efthimiou et al., 2009)

3.4 Creating metadata

Optionally, a sign language corpus may be enriched with metadata, i.e., with data describing the raw data, primary data, and (optional) secondary data. For example, information about the language of the raw data, the type of the annotations carried out, the availability of the corpus as a whole, etc. may be provided. Adding as much metadata as possible is advisable, as this increases the number of research questions for which the corpus can be consulted.

Metadata standards exist that provide a vocabulary of elements and attributes for describing metadata, the underlying objective being precisely that of standardization. Some standards contain vocabularies for the description of metadata only; others also provide an inventory of elements and attributes for the creation of secondary data, i.e., (linguistic) annotations. Among the former (metadata-only standards) are the Dublin Core Metadata Element Set (DCMES), the

Open Language Archives Community (OLAC) Metadata Standard, and the Isle Metadata Initiative (IMDI) Standard; among the latter (metadata and annotation standards) are the standard of the Text Encoding Initiative (TEI) and the Corpus Encoding Standard for XML (XCES).

To specify metadata in a sign language corpus, IMDI has proven useful. IMDI provides a hierarchical element structure with two types of metadata: *Catalogue metadata* describe the corpus as a whole, providing elements such as `Name`, `TitleID`, `Description`, `SubjectLanguage`, `DocumentLanguage`, `Location`, `Format`, or `Date`. *Session metadata* describe the individual sessions that make up the corpus, supplying elements like `Project`, `Content`, `Actors`, `Resources`, and `References`, which in turn may carry sub-elements. IMDI allows the extension of its vocabulary through key-value pairs. Such an extension has been created for sign language corpora, allowing for the provision of information on the hearing status, the sign language background, and the signing competence of the informants, the number of cameras used for the video recordings, etc. (Crasborn & Hanke, 2003). Figure 3.3 shows the use of key-value pairs of the IMDI extension for sign language corpora in the ECHO Corpus (Crasborn et al., 2007). IMDI was used to record metadata for the corpus of DSGS train announcements introduced in Section 3.6.

```
<Keys>
    <Key Name="Deafness.Status" Type="OpenVocabularyList">deaf</Key>
    <Key Name="Deafness.AidType" Type="OpenVocabularyList">none</Key>
    <Key Name="SignLanguageExperience.ExposureAge" Type="OpenVocabularyList">0</Key>
   <Key Name="SignLanguageExperience.AcquisitionLocation" Type="OpenVocabularyList">Stockholm</Key>
    <Key Name="SignLanguageExperience.SignTeaching" Type="OpenVocabularyList">good</Key>
    <Key Name="Family.Mother.Deafness" Type="OpenVocabularyList">deaf</Key>
   <Key Name="Family.Mother.PrimaryCommunicationForm" Type="OpenVocabularyList">sign language</Key>
    <Key Name="Family.Father.Deafness" Type="OpenVocabularyList">hard-of-hearing</Key>
   <Key Name="Family.Father.PrimaryCommunicationForm" Type="OpenVocabularyList">sign language</Key>
    <Key Name="Family.Partner.Deafness" Type="OpenVocabularyList">deaf</Key>
   <Key Name="Family.Partner.PrimaryCommunicationForm" Type="OpenVocabularyList">sign language</Key>
    <Key Name="Education.Age" Type="OpenVocabularyList"/>
    <Key Name="Education.SchoolType" Type="OpenVocabularyList"/>
    <Key Name="Education.ClassKind" Type="OpenVocabularyList"/>
    <Key Name="Education.EducationModel" Type="OpenVocabularyList">bilingual</Key>
    <Key Name="Education.Location" Type="OpenVocabularyList">Manillaskolan</Key>
    <Key Name="Education.BoardingSchool" Type="OpenVocabularyList"/>
    <Key Name="Handedness" Type="OpenVocabularyList">right</Key>
    <Key Name="Dialect" Type="OpenVocabularyList">Stockholm</Key>
    <Key Name="GeneralEducation.OccupationTrainedFor" Type="OpenVocabularyList"/>
    <Key Name="GeneralEducation.CurrentOccupation" Type="OpenVocabularyList">student</Key>
</Keys>
```

Figure 3.3: Key-value pair extension to IMDI for sign language corpora: Example from the ECHO Corpus (Crasborn et al., 2007)

3.5 Corpora used for automatic sign language processing

As previously mentioned, most corpora involving sign language so far were built with the aim of conducting linguistic analyses. Only few corpora have been created for the primary purpose of serving as data for automatic sign language processing systems, such as sign language recognition, sign language translation, or sign language animation systems. My work involved building such a corpus. More precisely, a parallel corpus for use in sign language machine translation and subsequent sign language animation was developed. In what follows, the type of data and sign language representation in previous parallel corpora used for sign language machine translation are discussed.

The Phoenix Parallel Corpus (Forster et al., 2012; Forster, Schmidt, Koller, Bellgardt, & Ney, 2014) is based on German weather reports interpreted into DGS and broadcast on the German TV station Phoenix. Figure 3.4 shows this setting. Transcriptions of the German speech were obtained through automatic speech recognition and manual postcorrection.[3] The broadcast videos served as raw data for the DGS side of the parallel corpus. The primary data on the DGS side consists of glosses and a very limited amount of non-manual information, e.g., about mouthings. The German and DGS sentences were aligned manually. Table 3.1 shows a German/DGS sentence pair. The corpus originally contained 3 000 sentence pairs (Forster et al., 2012) and was recently extended to 8 700 sentence pairs (Forster et al., 2014). It was used for building a statistical machine translation system as described in Section 4.3.1 and is presumably the largest corpus built specifically for this purpose.

Figure 3.4: Phoenix Parallel Corpus: German weather reports interpreted into DGS and broadcast on the German TV station Phoenix (figure from Stein et al., 2012)

[3] Stein, Schmidt, & Ney (2012) reported that the error rate of the speech recognition system was below 5%.

Die Temperaturen sinken in der Nacht auf 11 Grad an der Nordsee und 4 Grad an den Alpen. (‘At night, the temperatures fall to 11 degrees at the North Sea and 4 degrees near the Alps.’)	TEMPERATUR NACHT SINKEN 11 NORDEN SEE 4 GRAD ALPEN (‘TEMPERATURE NIGHT FALL 11 NORTH SEA 4 DEGREE ALPS’)

Table 3.1: Phoenix Parallel Corpus: German/DGS sentence pair (Stein et al., 2012)

The fact that the Phoenix corpus data stems from interpretation in a live setting has two implications: Firstly, since information was conveyed at high speed, the sign language interpreters omitted pieces of information from time to time. This led to an information mismatch between some German sentences and their DGS correspondences. The corpus creators therefore re-translated the DGS sentences into German, creating a second German version for each DGS sentence. Secondly, due to the high speed of transmission, the interpreters sometimes followed more closely the grammar of German than that of DGS in order to avoid memory buffer overload resulting from having to perform the additional cognitive task of reordering. In doing so, their output was more similar to Signed German (cf. Section 2.6) than DGS. However, Stein et al. (2012) maintain that this effect occurred only rarely.

Bungeroth et al. (2008) built the ATIS (Air Travel Information System) Corpus containing flight announcements in, among other languages, English, Irish Sign Language (ISL), German, and DGS. The corpus contains 595 sentences per language. Glosses were used as sign language representations. The ATIS Parallel Corpus was used in sign language machine translation experiments as described in Section 4.3.2.

3.6 Building a parallel corpus of German/DSGS train announcements for use in sign language machine translation and subsequent sign language animation

Chapter 1 introduced the project in the context of which this thesis is set. The application built as part of the project includes a statistical machine translation system that translates written German train announcements of the Swiss Federal Railways (*Schweizerische Bundesbahnen*) (SBB) into DSGS. Statistical machine translation systems require a parallel corpus as their training, development, and test data. Together with two Deaf bilingual researchers (one Deaf-of-Deaf, the other with a Deaf sibling), I therefore built a parallel corpus of German/DSGS train announcements as

data for my translation system. This is the first parallel corpus for use in automatic sign language processing to include DSGS. The corpus consists of 2 986 announcement pairs.

For the Phoenix Corpus introduced in Section 3.5, the sign language translations were already present in the form of videos at the time of the creation of the parallel corpus. Hence, the production of this corpus followed the order of steps outlined in Sections 3.1 and 3.2: Raw data (video recordings) was collected, after which primary data (segmentations and notations) was added. Conversely, for the ATIS Parallel Corpus (cf. Section 3.5) and the parallel corpus of train announcements described here, the translations into sign language (ISL/DGS and DSGS, respectively) had to be produced in a first step. Therefore, for our parallel corpus of German/DSGS train announcements, parts of the primary data (gloss and non-manual information notations) were produced first to obtain direct translations of the German originals. Following this, the raw data (video recordings) and the second part of the primary data (form notations) were created. The video recordings were required as a basis for the form notations.

Different from the Phoenix Corpus, a speech recognition step was not necessary to obtain the spoken language text: The German train announcements were provided to us in written form by the SBB. Hence, the process of translating 2 986 written German train announcements into DSGS consisted of the following steps:

1. Translating the written German train announcements into DSGS glosses and non-manual information;
2. Signing the announcements in front of a camera based on the glosses and non-manual information; and
3. Notating the form of both the manual and the non-manual components of the signing.

In what follows, these steps are discussed in more detail.

3.6.1 Translation into DSGS glosses and non-manual information

As stated in Chapter 2, DSGS is composed of five dialects. The project described here focused on the Zurich dialect. Hence, where several dialect variants of a sign existed, the one corresponding to the Zurich dialect was chosen. For example, there are two variants of the sign ZÜRICH (‘ZURICH’) in DSGS, one corresponding to the Zurich, Berne, Lucerne, and St Gallen dialects and the other to the Basel dialect.

We assigned glosses in a controlled way by referring to a DSGS lexicon and only introducing a new gloss if an appropriate one was not available. The DSGS lexicon we used has been under development since 1995 and currently contains about 9 000 signs, each represented by a gloss, a set of German keywords and a video clip for the citation form of the sign (Boyes Braem, 2001c). About half of the signs are notated in HamNoSys; these notations were used in the last step of creating the DSGS side of our parallel corpus (cf. Section 3.6.3). Together with its creator, I recently migrated the DSGS lexicon from its original (FileMaker) form to iLex, a sign language lexicon and corpus software (Hanke & Storz, 2008). Figure 3.5 shows the sample entry ZÜRICH_1A ('ZURICH_1A') from the DSGS lexicon in iLex.

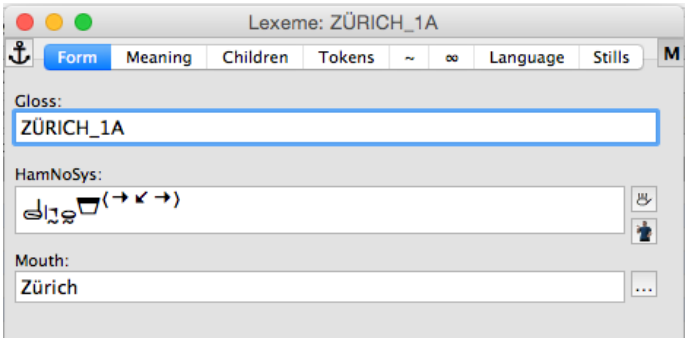


Figure 3.5: Entry in the DSGS lexicon in the iLex software

For several concepts that were specific to the train announcement domain, no signs existed in the DSGS lexicon. Analogous to what is done in such cases for DSGS interpreting on Swiss television and to what had been done for an online lexicon of technical signs (Boyes Braem, Groeber, Stocker, & Tissi, 2012), we asked an expert group of DSGS signers to discuss appropriate signs for these concepts, e.g., for *Betriebslagemonitor* ('operational status monitor'), *Buskante* ('bus edge'), or *Fahrleitungsstörung* ('overhead line disruption'). 18 signs were newly coined in such a way.

For train names, we introduced the following distinction: If a commonly used spoken language abbreviation for a train name existed, such as IC for *InterCity*, we fingerspelled (cf. Section 2.5) the letters of the abbreviation, adding an outward stamping movement after each fingerspelling sign as is common when signing abbreviations. In all other cases, we used existing DSGS lexical signs. Often, we combined two or more existing signs, e.g., EURO ('EURO') and NACHT ('NIGHT') for *EuroNight*, STADT ('CITY'), NACHT ('NIGHT'), and LINIE ('LINE') for *CityNightLine*, or NACHT ('NIGHT') and VOGEL ('BIRD') for *Nightbird*.

To obtain information about signs for places with train stations in Switzerland, we released a call via a web platform popular among the DSGS community,[4] asking Deaf DSGS signers to send us videos of signs for places well known to them. In addition, as part of a research module, students of the DSGS interpreting training programme at the University of Applied Sciences of Special Needs Education Zurich attended events organized by the Deaf community in the German-speaking part of Switzerland, where they asked Deaf persons for signs for places. In this way, we collected 531 videos of signs for 284 places, such as *Allschwil*, *Spreitenbach*, or *Waltensburg*. The two Deaf collaborators analyzed the contributions. In total, of the 590 place names in our DSGS train announcements, 320 received lexical signs, of which parts were from the DSGS lexicon mentioned earlier and parts from the collected data just described. A more comprehensive data collection process could have yielded further lexical signs.

The remaining place names were fingerspelled. When fingerspelling place names starting with *St.* or *San* (as in *St. Margarethen* or *San Giovanni*), a dot was signed after *St* and a pause after *San*, like so: S-T DOT M-A-R-G-A-R-E-T-H-E-N and S-A-N [pause] G-I-O-V-A-N-N-I. With place names that consisted of two parts (such as *Hüntwangen-Wil*), we explicitly added the sign STRICH ('DASH') inbetween. In some cases, we used combinations of lexical signs and fingerspellings, such as for *Schinznach Bad*, where we fingerspelled SCH-I-N-Z-N-A-CH[5] and appended the lexical sign BAD ('BATH'). We did the same for place names involving specifications of orientations such as NORD ('NORTH'), SÜD ('SOUTH'), WEST ('WEST'), and OST ('EAST'). Similarly, specifications like *am*, *bei*, or *an der* ('at/close to (the)') were expressed through the sign NAHE ('NEAR'), e.g., *Beinwil am See* was translated into DSGS as B-E-I-N-W-I-L NAHE SEE ('B-E-I-N-W-I-L NEAR LAKE').

The German polite form *Sie* appeared frequently in the spoken language side of our train announcements, as in the following two examples: *Bitte folgen Sie den Wegweisern.* ('Please follow the posted signs.') and *Bitte warten Sie im Sektor A.* ('Please wait in Sector A.'). This polite form does not exist in DSGS. We therefore provided DSGS translations for the German announcements *Bitte den Wegweisern folgen* ('Please follow the posted signs') and *Bitte im Sektor A warten* ('Please wait in Sector A').[6]

[4] `http://www.deafzone.ch/` (last accessed October 1, 2015).

[5] Recall from Section 2.5 that the DSGS finger alphabet features dedicated signs for -CH- and -SCH-.

[6] The resulting translations are the same as for the German announcements *Bitte folge den Wegweisern* and *Bitte warte im Sektor A*, respectively.

We defined the following sign string format for time specifications: <STUNDEN> UHR <MINUTEN> (‘<HOUR NUMBER> CLOCK <MINUTE NUMBER>’), e.g., SIEBZEHN UHR FÜNFZEHN (‘SEVENTEEN CLOCK FIFTEEN’). This format was the result of a focus group study carried out with Deaf signers, as reported in Section 5.3.3. The format initially chosen was that of a timetable, UHR <STUNDEN> PUNKT <MINUTEN> (‘CLOCK <HOUR-NUMBER> DOT <MINUTE-NUMBER>’), e.g., UHR SIEBZEHN PUNKT FÜNFZEHN (‘CLOCK SEVENTEEN DOT FIFTEEN’). However, the participants of the focus group preferred the format more familiar to them, <STUNDEN> UHR <MINUTEN>.

Recall that from Section 2.1 that non-manual components make up an important part of signing. In addition to the glosses, we therefore created non-manual information notations. We included information about head movement, movement of eyebrows, eye aperture, and mouthings. Mouthings are not based on a closed-class vocabulary. For the other types of non-manual components, we introduced the following possible values, departing from the annotation scheme of Neidle (2002, 2007):

- Head: neutral, forward, forward/shake (simultaneously), back, back/right (simultaneously), back/left (simultaneously), nod, circular
- Eyebrows: neutral, up, furrowed
- Eye aperture: neutral, wide

Table 3.2 shows the gloss and non-manual information notation of the DSGS translation of the German announcement *Wir werden Sie weiter informieren.* (‘We will keep you informed.’). Note that the boundaries of the non-manual components align with those of manual activities. This is because the non-manual components in our corpus serve linguistic rather than affective functions (cf. Section 2.1).

Gloss **Gloss: translation**	DANN ‘THEN’	WEITER ‘FURTHER’	INFORMIEREN ‘INFORM’
Head	down	nod	neutral
Eyebrows	up		
Eye aperture	neutral	wide	
Mouthing	/dann/	/weiter/	/informieren/

Table 3.2: Non-manual information for the DSGS translation of the German train announcement *Wir werden Sie weiter informieren.* (‘We will keep you informed.’)

3.6.2 Video recording

The signing was recorded in the studio of the Swiss Deaf Association. One camera was used. This was deemed sufficient at the time, as the sole purpose of the video recordings was to serve as basis for the subsequent form notation step. However, in retrospect, using a second camera to record the signing as a basis for the subsequent form notation step might have been beneficial, as it would have allowed for an even more accurate description of the form of signs in cases where there is contact between two parts of the body (e.g., between the two hands, between one hand and the upper body, etc.).

3.6.3 Form notation

Applying statistical machine translation implies that the translation output is made up of segments (n-grams) from the target side of the parallel corpus. In our overall application, the machine translation step was succeeded by a sign language animation step (cf. Chapter 1). To provide sufficient information for the animation step, every sign in the DSGS side of the parallel corpus needed to be associated with a form description from which motion data (i.e., frame-based information about the rotations of joints and about the morphs in the face; cf. Section 5.3.1) could be generated during the animation step.

The animation system used for generating synthesized DSGS train announcements[7] accepts information in an XML representation of HamNoSys, the Signing Gesture Markup Language (SiGML) (Elliott, Glauert, Kennaway, & Marshall, 2000). Information about the form of the manual activities in the DSGS side of our parallel corpus was therefore encoded in HamNoSys. Where possible, we relied on HamNoSys notations from the DSGS lexicon of Boyes Braem (2001c). The notations not available in the lexicon were created by one of the two Deaf DSGS signers in our project. Each notation was additionally checked by a Deaf expert at the University of Hamburg, the place of origin of HamNoSys.

The notations reflected the citation forms of the signs, i.e., their forms represented in the lexicon and not their forms in a given context. Ideally, each occurrence of a sign in a signed utterance would be notated individually, taking account of possible coarticulation effects. For example,

[7] The system is described in Section 5.3.1.

in our DSGS train announcements, the starting location of one of the occurrences of the sign AUSFALL (‘CANCELLATION’) is slightly changed due to the preceding sign ZOLLIKOFEN (place name), which is performed to the signer’s right following an indexical (pointing) at that location. The starting location of this occurrence of the sign AUSFALL is shown in Figure 3.6 (left) in contrast to the starting location of the citation form of the sign (right). Ideally, separate HamNoSys notations (containing different location symbols) would be introduced for these two forms. However, such a notation procedure was outside of the scope of this thesis. Therefore, we created only one HamNoSys notation for each sign, using as reference the form of the sign in the DSGS lexicon.

Figure 3.6: Coarticulation effect: Occurrence of the sign AUSFALL (‘CANCELLATION’) in the DSGS train announcements (left) vs. citation form in the DSGS lexicon (Boyes Braem, 2001c) (right)

The animation system used to synthesize the DSGS announcements has its own mechanism for interpolating between signs. Hence, only the signs and not the inter-sign transitions were notated in HamNoSys. This corresponds to narrow segmentation (cf. Section 3.2.1). A sample announcement notated in HamNoSys is shown in Table 3.3.

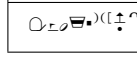
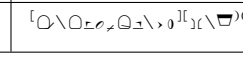
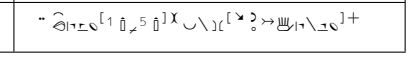
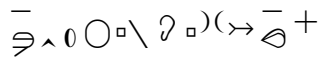
DANN ‘THEN’	WEITER ‘FURTHER’	INFORMIEREN ‘INFORM’
		

Table 3.3: Glosses and HamNoSys notations for the DSGS translation of the German train announcement *Wir werden Sie weiter informieren.* (‘We will keep you informed.’)

SiGML, the XML representation of HamNoSys, has two variants: HNS SiGML, which is essentially a list of HamNoSys symbols in XML form, and Gestural SiGML, which is more suitable for automatic processing and supports some extensions to the HamNoSys model. Gestural SiGML is the variant used for the DSGS avatar. Figure 3.7 shows the HamNoSys notation of the DSGS sign LAUTSPRECHER (‘LOUDSPEAKER’) along with the corresponding Gestural SiGML code for the manual activity of the sign. The sign is performed by opening and closing the dominant hand next to the ear.



```
<sign_manual>
  <handconfig ceeopening="slack" hand shape="ceeall"
  mainbend="bent"/>
  <handconfig extfidir="u"/>
  <handconfig palmor="l"/>
  <location_bodyarm contact="close" location="head"
  second_location="ear" second_side="right_beside"
  side="right_beside"/>
  <rpt_motion repetition="fromstart">
      <tgt_motion>
      <changeposture/>
      <handconfig hand shape="pinchall" mainbend="bent"/>
      </tgt_motion>
  </rpt_motion>
</sign_manual>
```

Figure 3.7: HamNoSys notation and corresponding SiGML code for the manual activity of the DSGS sign LAUTSPRECHER (‘LOUDSPEAKER’)

Just like information about the manual activities, information about the non-manual components also needed to be specified not just with descriptive labels (such as “eyebrows raised” or “head nod”) but with precise information about the physical form of the non-manual behavior, from which motion data could subsequently be created. No HamNoSys symbols exist for encoding non-manual aspects of signing; instead, this information is specified in SiGML directly. SiGML provides an inventory of attribute values for expressing non-manual information. Examples of such values are RB for “(both) eyebrows raised” or NO for “head nod” (Hanke, 2001). However, this inventory was not sufficient to cover the non-manual information present in our DSGS train announcements. I therefore extended it together with the two Deaf project members. Moreover, I modified the geometric descriptions of existing values. This work is discussed in more detail in Section 5.3.2. I then mapped the descriptive non-manual information labels to SiGML values. Table 3.4 shows this mapping for the non-manual features of the previously introduced announcement example.

Gloss **Gloss: translation**	DANN ‘THEN’	WEITER ‘FURTHER’	INFORMIEREN ‘INFORM’
Head	down → N2	nod → NO	neutral
Eyebrows	up → RB		
Eye aperture	neutral	wide → WB	
Mouthing	/dann/	/weiter/	/informieren/

Table 3.4: Mapping between descriptive labels and SiGML values for the non-manual features of the DSGS translation of the German train announcement *Wir werden Sie weiter informieren.* (‘We will keep you informed.’)

JASigning accepts mouthing information (cf. Section 2.1) transcribed in the Speech Assessment Methods Phonetic Alphabet (SAMPA) (Wells, 1997), a machine-readable version of the International Phonetic Alphabet (IPA). As an example, the DSGS mouthing /Lautsprecher/ ('loudspeaker') is notated in SAMPA as `'laUt|,SprE|C@r|`.[8] The Bonn Machine-Readable Pronunciation Dictionary (BOMP) (Portele, Krämer, & Stock, 1995) provides 141 230 SAMPA notations for German. I used these notations and modified them, as sign language mouthings often correspond to only a portion of a spoken language word (cf. Section 2.1). For example, I changed the BOMP notation `?Ent|'SUl|dI|gUN|` to `EntSUldUN`. Missing notations were also added.

SAMPA notations can be embedded into SiGML code as attribute values. Figure 3.8 shows the SiGML code for the manual activity and the mouthing of the DSGS sign LAUTSPRECHER ('LOUDSPEAKER'). Non-manual information is given inside a `<sign_nonmanual>` element. The SAMPA transcription of /Lautsprecher/ is specified inside a `<mouth_picture>` element. The code for the manual activity is provided inside a `<sign_manual>` element (cf. Figure 3.7).

```
<hamgestural_sign gloss="LAUTSPRECHER">
  <sign_nonmanual>
    <mouthing_tier>
      <mouth_picture picture="laUtSprEC@r"/>
    </mouthing_tier>
  </sign_nonmanual>
  <sign_manual>
    <handconfig ceeopening="slack" hand shape="ceeall"
    mainbend="bent"/>
    <handconfig extfidir="u"/>
    <handconfig palmor="l"/>
    <location_bodyarm contact="close" location="head"
    second_location="ear"
    second_side="right_beside" side="right_beside"/>
    <rpt_motion repetition="fromstart">
      <tgt_motion>
        <changeposture/>
        <handconfig hand shape="pinchall" mainbend="bent"/>
      </tgt_motion>
    </rpt_motion>
  </sign_manual>
</hamgestural_sign>
```

Figure 3.8: SiGML code for the manual activity and mouthing of the DSGS sign LAUTSPRECHER ('LOUDSPEAKER')

Gestural SiGML allows for a fine-grained specification of entire signs, the manual components, and the non-manual components of a sign well beyond the simple example shown in

[8] Syllable and accent information are ignored in the avatar system used, as such information makes little difference to the visual appearance. Hence, the avatar system essentially reads the SAMPA string shown here as `laUtSprEC@r`.

Figure 3.8. Each sign (represented as a <hamgestural_sign> element) may carry three attributes: duration, speed, and timescale. In addition, each non-manual tier element (such as <mouthing_tier> shown in Figure 3.8) may contain a child element <…_par> (e.g., <mouthing_par>) that causes the non-manual features embedded in it to be executed in parallel rather than in sequence.[9] Each non-manual tier element may also carry an attribute presynchronization or postsynchronization to control the synchronization of the non-manual features within it. An attribute fitpicturetomanual can be specified for the <mouthing_tier> element to synchronize the duration of the mouthing and the manual activity of a sign. A mouthing can also be held or stretched over multiple signs with the <mouth_meta> element. Similarly, the <hamgestural_segment> element allows non-manual features to be applied to multiple signs. Figure 3.9 displays schematic HNS SiGML and Gestural SiGML code. Underlined are the elements and attributes described that are exclusive to Gestural SiGML. As stated above, Gestural SiGML was the variant used for the DSGS avatar.

While all of the above features are part of the Gestural SiGML document type definition (DTD),[10] not all of them have actually been implemented in the avatar system used to synthesize the train announcements. Finding ways of achieving the effects of these features nevertheless (through workarounds) presented a challenge. This work is reported in Section 5.3.2.

In summary, equipping the sign language side of our parallel corpus with all information necessary for the animation step that forms part of the overall process of automatically translating written German train announcements into synthesized DSGS consisted of

- linking glosses to HamNoSys notations, which could then be converted to SiGML elements;
- linking non-manual information notations to SiGML attribute values; and
- linking mouthing information to SAMPA transcriptions, which could then be embedded in SiGML code.

Figure 3.10 visualizes these linkings for the previously introduced sentence *Wir werden Sie weiter informieren.* (‘We will keep you informed.’).

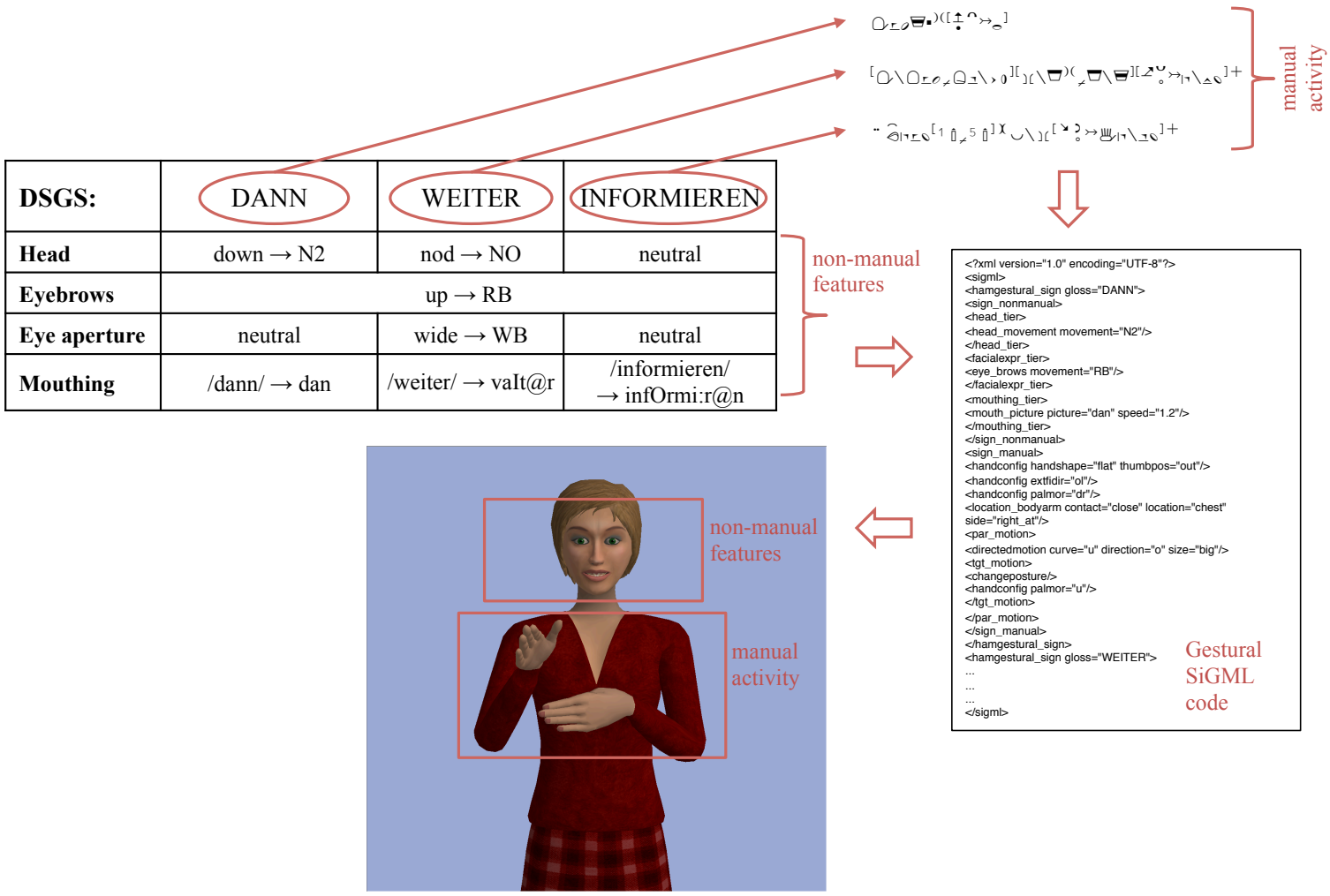
[9] Apart from the <mouthing_tier> element shown in Figure 3.8, Gestural SiGML permits the elements <facialexpr_tier>, <shoulder_tier>, <body_tier>, <head_tier>, <eye gaze_tier>, and <extra_tier>.

[10] The DTD is available at http://www.visicast.cmp.uea.ac.uk/sigml/sigml.dtd (last accessed October 1, 2015).

```
<sigml>
  <hns_sign>
    <hamnosys_nonmanual>
      <hnm_shoulder tag=""/>
      <hnm_body tag=""/>
      <hnm_head tag=""/>
      <hnm_eye gaze tag=""/>
      <hnm_eyebrows tag=""/>
      <hnm_eyelids tag=""/>
      <hnm_nose tag=""/>
    <hnm_mouthpicture picture=""/>
     <hnm_mouthgesture tag=""/>
    <hnm_extramovement tag=""/>
    </hamnosys_nonmanual
    <hamnosys_manual>
      ...
    </hamnosys_manual>
  </hns_sign>
</sigml>
```

```
<sigml>
  <hamgestural_segment>
   <hamgestural_sign duration="" speed="" timescale="">
      <sign_nonmanual>
     <shoulder_tier presynchronization="slight_delay|start_slightly_ahead"
     postsynchronization="lasts_longer|ends_before">
      <shoulder_par> // available for all non-manual tier elements
       <shoulder_movement movement=""/>
        </shoulder_par>
      </shoulder_tier>
      <body_tier>
     <body_movement movement=""/>
      </body_tier>
      <head_tier>
     <head_movement movement=""/>
     <avatar_morph movement="HPSF" amount="2.0" timing="x m t m s l x"/>
      </head_tier>
      <eye gaze_tier>
     <eye_gaze direction=""/>
      </eye gaze_tier>
      <facialexpr_tier>
     <eye_brows movement=""/>
      <eye_lids movement=""/>
        <nose movement=""/>
      </facialexpr_tier>
     <mouthing_tier fitpicturetomanual="true|false">
      <mouth_picture picture=""/>
      <mouth_gesture movement=""/>
      </mouthing_tier>
      <extra_tier>
     <extra_movement movement=""/>
      </extra_tier>
    </sign_nonmanual>
    <sign_manual>
      ...
     <tgt_motion duration="" speed="" timescale="">
      <directedmotion direction="o" size="small"/>
      <handconstellation contact="touch"/>
      </tgt_motion>
      ...
    </sign_manual>
   </hamgestural_sign>
  </hamgestural_segment>
</sigml>
```

Figure 3.9: Comparison between schematic HNS SiGML (left) and Gestural SiGML (right)



DSGS:	DANN	WEITER	INFORMIEREN
Head	down → N2	nod → NO	neutral
Eyebrows	up → RB		
Eye aperture	neutral	wide → WB	neutral
Mouthing	/dann/ → dan	/weiter/ → vaIt@r	/informieren/ → infOrmi:r@n

Figure 3.10: Equipping the sign language side of the parallel corpus with information required for the animation step

3.6.4 Corpus profile

The resulting parallel corpus of German/DSGS train announcements consists of 2 986 announcement pairs. The average announcement length in tokens is approximately 14 for the German side and 15 for the DSGS side. I randomly divided the data into ten folds. For the machine translation experiments reported in Chapter 4, I used folds 1 to 8 for the training set, fold 9 for the development set, and fold 10 for the test set. Table 3.5 shows the profile of each of these sets. On the sign language side, "types" refers to the number of distinct glosses (the vocabulary), while "tokens" denotes the sum of all individual gloss occurrences in the corpus. On both sides, "singletons" is the percentage of types that occur only once (i.e., are *hapax legomena*). "Out-of-vocabulary units" refers to the number of types that appear in the test set but not in the training set.

	German	**DSGS**
Training set		
Announcements	← 2 393 →	
Tokens	33 342	35 524
Types	848	871
Singletons	264	261
Development set		
Announcements	← 293 →	
Tokens	4 012	4 265
Types	448	458
Singletons	167	173
Test set		
Announcements	← 300 →	
Tokens	4 208	4 503
Types	408	417
Singletons	141	143
Out-of-vocabulary items	25	28

Table 3.5: Parallel corpus of train announcements: Training, development, and test set

For comparison, the profiles of the 2012 version of the Phoenix Parallel Corpus and the ATIS Parallel Corpus (cf. Section 3.5) are given in Table 3.6.[11] If the training set of a parallel corpus has a high type-token ratio (i.e., a high number of tokens per type) and a low singleton rate, its size is taken to be "big enough for a given domain to train a translation system with decent quality"

[11] The bulk of the machine translation experiments on the Phoenix Corpus described in Chapter 4 are based on the 2012 version of the corpus, not the later (2014) version. The numbers given in Stein et al. (2012) for the 2012 version deviate slightly from those in Forster et al. (2012) but are on the same order of magnitude.

(Stein et al., 2012, p. 337). The training set of our parallel corpus of train announcements has a type-token ratio of 39 (German) and 40 (DSGS). The corresponding numbers for the Phoenix Corpus are 23 (German) and 30 (DGS); for the ATIS Corpus, 10 (English) and 11 (ISL). The singleton rate for the training set of our corpus is 31% (German) and 30% (DSGS); for the Phoenix Corpus, it is 36% for both German and DGS; for the ATIS Corpus, 33% (English) and 27% (ISL). To summarize, our parallel corpus of train announcements has a higher type-token ratio than both the Phoenix Corpus and the ATIS Corpus and a lower singleton rate than the Phoenix Corpus.

	German	DGS
Training set		
Sentences	← 2 565 →	
Tokens	41 306	31 208
Types	1 763	1 027
Singletons	641	371
Test set		
Sentences	← 512 →	
Tokens	8 230	6 115
Types	915	570
OOV	133	86

	English	ISL
Training set		
Sentences	← 418 →	
Tokens	3 008	3 028
Types	292	265
Singletons	97	71
Development set		
Sentences	← 59 →	
Tokens	429	431
Types	134	131
Test set		
Sentences	← 118 →	
Tokens	999	874
Types	174	148

Table 3.6: 2012 Phoenix Parallel Corpus (left) and ATIS Parallel Corpus (right): Profiles

3.7 Summary

This chapter has dealt with the process of creating a sign language corpus, addressing one of the research questions of this thesis as outlined in Chapter 1. In its default form, the process consists of obtaining raw data (collecting existing or producing new video recordings of signing), creating primary data (transcription, i.e., segmentation and notation), and possibly adding metadata and secondary data. I have shown that the order of steps is different in the case of a parallel corpus for which the sign language data does not yet exist. This was the case for a parallel corpus of German/DSGS train announcements built by myself in collaboration with two Deaf DSGS signers. As an initial step, the DSGS translations were produced, corresponding to parts of the primary data (gloss and non-manual information notations). Subsequently, the raw data (video

recordings) was created, and only then was the second part of the primary data (form notations) produced. The video recordings were necessary to have as a basis for the form notations.

The German announcements were available in written form. For the gloss and non-manual information notations, conventions were developed to ensure consistency. This included, for example, introducing two different ways of signing train names. Signs for concepts that were specific to the train announcement domain were added. Signs for places with train stations in Switzerland were collected as part of a crowdsourcing approach.

The non-manual information in the corpus consisted of information about head movement, movement of eyebrows, eye aperture, and mouthings. In the future, additional features, such as shoulder movements or eye blink, might be included to further increase the acceptance of the resulting sign language animations (sign language animation acceptance is discussed in Chapter 5).

Within the overall application of our project, the machine translation step is succeeded by a sign language animation step. Since the input to the animation system consists of segments from the target side of the parallel corpus recombined by the machine translation system, it was necessary to include all information required for the animation step in the DSGS side of the parallel corpus. Most importantly, this meant providing form descriptions of both the manual and the non-manual activities present in the DSGS train announcements, i.e., linking glosses to HamNoSys notations, non-manual information notations to SiGML attribute values, and mouthing information to SAMPA transcriptions. I have shown that we created one HamNoSys notation for each sign type. Ideally, a separate notation would be produced for each occurrence of a sign in context (i.e., each sign token), which would allow for taking account of possible coarticulation effects. However, this is a very time-consuming task.

The SiGML inventory of non-manual information was not sufficient for our purposes and had to be extended. Moreover, existing geometric descriptions had to be modified. Not all SiGML features have actually been implemented in the avatar system used to synthesize the train announcements. How this was dealt with is addressed in Chapter 5.

The resulting parallel corpus of German/DSGS train announcements holds 2 986 announcement pairs. Hence, its size is comparable to the initial size of the Phoenix Parallel Corpus, and it is substantially larger than the ATIS Parallel Corpus. Our parallel corpus of train announcements has a higher type-token ratio than both the Phoenix Corpus and the ATIS Corpus and a lower

singleton rate than the Phoenix Corpus, both of which are desirable properties for data used in machine translation.

Chapter 4

Sign language machine translation

[This chapter is an extension of Ebling (2010) and Ebling (2013).]

4.1 Paradigms

The two major paradigms of machine translation are rule-based and statistical machine translation (SMT). SMT is a *data-driven* (or, *corpus-based*) paradigm in that it requires a sententially aligned bilingual corpus, a *parallel corpus* (or, *bitext*) (cf. Chapter 3). At the very least, the corpus is divided into a training set and a smaller test set. Typically, there is a third set, the development set. It is used to tune the parameters of a system.

SMT started out as word-based translation (Brown et al., 1990) and gradually evolved into phrase-based translation (Koehn, Och, & Marcu, 2003). Extensions of phrase-based translation include hierarchical (Chiang, 2005) and factored (Hoang, 2007) SMT.

Sign language machine translation refers to the process of automatically translating from a spoken language into a sign language, from a sign language into a spoken language, or from one sign language into another. Research on sign language machine translation started in the 1990s. At that time, the rule-based paradigm dominated, and grammar formalisms such as Tree-Adjoining Grammar (TAG) or Head-Driven Phrase Structure Grammar (HPSG) were applied.[1] Today, sign language machine translation takes place mostly within the statistical paradigm. My work

[1] For an overview of earlier rule-based sign language machine translation systems, see Huenerfauth (2003).

in translating German train announcements into Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS), as reported in Section 4.4, relied on SMT as well.

4.2 Evaluation

Machine translation evaluation is the process of assessing the quality of the output of a machine translation system, the *candidate translation* (or, *hypothesis*). Evaluation may be performed either by a human (*human evaluation*) or by a machine (*automatic evaluation*). With automatic evaluation, the output of the translation system is compared against one or multiple *reference translations* using a metric.

The most common metrics for automatic evaluation are either distance-based or n-gram-based (Estrella, 2008). Distance-based metrics compute the minimum number of edit operations (substitutions, insertions, and deletions) required to transform a candidate translation into a reference translation. For example, the Word Error Rate (WER) (Tillmann, Vogel, Ney, Zubiaga, & Sawaf, 1997), an evaluation metric from speech recognition, calculates edit distance based on tokens. The final score is computed by dividing the sum of all necessary edit operations by the number of tokens in the candidate translation. The metric has two shortcomings: Firstly, all tokens receive the same weight, i.e., there is no distinction between deleting, e.g., a punctuation symbol and a content word. Secondly, since candidate/reference token pairs are compared sequentially, the metric does not allow for variation in word order. The Position-Independent Word Error Rate (PER) (Nießen, Och, Leusch, & Ney, 2000) was introduced to overcome this deficiency. PER treats the candidate and reference translations as bags of words and thus abstracts over position.

A more recent distance-based evaluation metric is the Translation Edit Rate (TER) (Snover, Dorr, Schwartz, Micciulla, & Makhoul, 2006). It differs from WER in that it allows for shifts of tokens/phrases in addition to the basic edit operations. A drawback unique to this metric is that neither the length of the token sequences that are shifted nor the distance across which they are shifted are taken into consideration. TER also inherits the shortcomings of WER, i.e., punctuation marks are treated as regular tokens, and all operations have a uniform cost of 1. In addition, case corrections (lowercase to uppercase or vice versa) count as regular edit operations.

N-gram-based metrics are the most widely used automatic evaluation metrics. Among these, the Bilingual Evaluation Understudy (BLEU) metric (Papineni, Roukos, Ward, & Zhu, 2002) is most frequently applied. BLEU is based on n-gram precision. In its basic form, n-gram precision

is computed as the number of correctly translated word n-grams divided by the total number of word n-grams in the candidate translation. The number of correctly translated word n-grams is equal to the number of word n-grams in the candidate translation that appear in the reference translation(s). The problem with basic n-gram precision is that a candidate translation that contains only one word will receive a precision of 1 if one of the reference translations contains this word as well. BLEU therefore includes a modified n-gram precision score: The number of possible n-gram matches is limited to the number of occurrences of this n-gram in a single reference translation. In other words, for each n-gram in a candidate translation, its number of occurrences in each of the reference translations is determined, the maximum value is chosen and divided by the total number of n-grams in the candidate translation. Modified n-gram precision is calculated separately for each n-gram order. The n-gram order in BLEU usually ranges from 1 to 4. N-grams of higher order to some extent capture grammatical well-formedness. However, this is not to say that BLEU takes into account syntactic structure explicitly.

As a result of computing n-gram precision, BLEU automatically penalizes candidate translations that are longer than their reference translations. To penalize candidate translations that are shorter than their reference translations, a brevity penalty score (BP) was introduced. BP is computed over the entire corpus.[2] It is defined as in Equation 4.1, where c is the length of the candidate translation and r the length of the reference corpus.

$$BP = \begin{cases} 1 \text{ if } c > r \\ e^{(1 - \frac{r}{c})} \text{ if } c \leq r \end{cases} \tag{4.1}$$

The overall BLEU score is computed as the geometric mean of the modified n-gram precisions, p_n, multiplied by the exponential brevity penalty score, BP, as shown in Equation 4.2. N is the maximum n-gram length and w_n a positive weight (the weights together sum up to 1).

$$BLEU = BP \cdot exp(\sum_{n=1}^{N} w_n \, log \, p_n) \tag{4.2}$$

Equation 4.2 shows that the BLEU score is 0 if any of its factors is 0. This is just one of several

[2] Papineni et al. (2002) posited that computing it over individual sentences would lead to a penalty that is too severe.

shortcomings of this metric. BLEU has also been shown to severely penalize only small differences in length between a candidate and a reference translation. Moreover, Callison-Burch, Osborne, & Koehn (2006) point out that there are often "millions of variations on a hypothesis translation that receive the same Bleu score" (p. 1), stressing that "[a]s the number of identically scored variants goes up, the likelihood that they would all be judged equally plausible goes down" (p. 4).

Some drawbacks of BLEU were tackled in a metric developed by the National Institute of Standards and Technology (NIST) and referred to by that name (Doddington, 2002). NIST applies the arithmetic rather than the geometric mean, thus providing a relief from the problem of total zero-value scores. Furthermore, NIST includes a modified brevity penalty score that penalizes small differences in length between a candidate and a reference translation less severely. Its major conceptual improvement is the introduction of n-gram weights: Less frequent n-grams are assigned a higher weight than more frequent ones, as they are considered to be more informative. A drawback unique to NIST is that its score increases with the amount of text that is used for the evaluation. This means that the metric contains no upper bound, which makes comparisons of NIST scores obtained with different amounts of data essentially impossible.

The machine translation approaches dealt with in the following sections were evaluated using combinations of WER, PER, TER, BLEU, and NIST. Note that WER, PER, and TER are error measures, which means a lower score is indicative of higher translation quality, while BLEU and NIST are quality measures, which means a higher score is indicative of higher translation quality.

4.3 Limited-domain statistical sign language machine translation

SMT systems require large amounts of training data. The consensus is that "more data are better data" (Mercer, 1993, p. 18), with a more recent restriction according to which it is particularly desirable to have in-domain data. Large parallel corpora for spoken languages, e.g., the European Parliament Proceedings (Europarl) Parallel Corpus (Koehn, 2005), range on the order of 50 million words per language.[3] For sign languages, presumably the largest parallel corpus built for use in machine translation is the Phoenix Corpus described in Chapter 3, which contains 8 700

[3] `http://www.statmt.org/europarl/` (last accessed September 30, 2015).

sentence pairs (Forster et al., 2014). As has been shown, compiling a parallel corpus that involves sign language is a heavily time-consuming task even if the sign language representation consists of glosses only, as is true for the Phoenix Corpus. SMT systems that are trained on such comparatively small data sets can be expected to work well only if they operate on restricted domains. Such is the case for an SMT system that operates on the Phoenix Corpus, translating weather reports from German into German Sign Language (*Deutsche Gebärdensprache*) (DGS) and vice versa. Another limited-domain statistical sign language machine translation system translates air travel information between English and Irish Sign Language (ISL), German and ISL, English and DGS, and German and DGS. In what follows, the two systems are described in more detail. The research findings most relevant for my own work (cf. Section 4.4) are presented.

4.3.1 Weather reports

Stein et al. (2012) translated weather reports from German to DGS and vice versa, using the Phoenix Parallel Corpus described in Chapter 3. The researchers applied two in-house translation systems: a phrase-based SMT system, PBT (Zens & Ney, 2008), and a hierarchical phrase-based SMT system, JANE (Vilar, Stein, Huck, & Ney, 2010).

4.3.1.1 Translation from DGS to German

The system translating from DGS to German was part of a larger application that also included automatic sign language recognition, instantiating the pipeline of type b) described in Chapter 1 (cf. Figure 1.1). To show that DGS and German are indeed sufficiently different to warrant a translation step between them (in addition to the translation step being justified by the mere fact that DGS and German are two separate languages), Stein et al. (2012) performed a sanity check: Instead of providing the German translation output as hypothesis, they used a lowercased version of the source side of the test set (DGS gloss text) and compared it with the German reference (i.e., the target side of the test set). This resulted in the following data being used for the evaluation:

- **Input:** DGS (source side of test set)
- **Hypothesis:** DGS (source side of test set), lowercased
- **Reference:** German (target side of test set), lowercased

For my own work (cf. Section 4.4), I applied a similar configuration not as a sanity check but as a baseline. The results of the sanity check experiment by Stein et al. (2012) are given in the first row of Table 4.1: The BLEU score was 2.6, in comparison to 22.0 (second row) when using the German output of the JANE translation system as hypothesis. This difference of almost 20 BLEU points indicates that a translation step is, indeed, in order.

System	**BLEU**	**TER**	**PER**
Sanity check	2.6	81.1	74.8
JANE output	22.0	74.0	65.1

Table 4.1: Translation from DGS to German: Sanity check

Stein et al. (2012) further improved their system by applying alternative optimization methods: When the size of the parallel corpus was still on the order of 3 000 sentences (cf. Chapter 3), reserving a few hundred sentences for the development set was likely to decrease translation performance, as those sentences were lost as training data for the system. Therefore, the researchers split the training data into five subsets of 513 sentences each and trained a separate system for each subset. In each optimization iteration, they merged the n-best lists of the individual systems to obtain a full translation of the training set. This yielded results that were significantly higher than the baseline system, which consisted of a traditional split into a training set of around 2 000 sentences and a development and test set of about 500 sentences each. The results are given in Table 4.2: The first row shows the performance of the baseline system, the second the performance of the system using the alternative optimization method ("leave-513-out").

System	**BLEU**	**TER**
Baseline (traditional split training/dev)	22.0	73.9
Leave-513-out	23.0	72.0

Table 4.2: Translation from DGS to German: Applying the JANE system with an alternative optimization method

Another notable experiment for the translation direction DGS to German consisted of remedying the fact that the sign language interpreters, when interpreting the weather reports into DGS, tended to leave out information from the German original. As reported in Chapter 3, Forster et al. (2012) therefore created additional German references that were direct back-translations of the DGS gloss text. Table 4.3 shows the results for using the original transcripts of the German

speech (first row) versus using both the original transcripts and the newly created German translations of the DGS gloss text (second row). The BLEU scores improved by 7.0 as a result of providing the additional references.[4]

System	BLEU	TER
One reference (original transcripts of German speech)	31.8	61.8
Two references (original transcripts and translations of gloss text)	38.8	53.9

Table 4.3: Translation from DGS to German: Using two German references instead of one

Lastly, Forster et al. (2014) reported the results obtained from using the extended version of the parallel corpus, amounting to 8 700 sentence pairs (compared to originally 3 000 sentences), as data for the translation system. Table 4.4 provides the results for these experiments, showing an increase of 1.6 BLEU (single reference: 31.8 to 33.4) and 2.1 BLEU (two references: 38.8 to 40.9), respectively, compared to the results displayed in Table 4.3.

System	BLEU	TER
Extended parallel corpus, one reference	33.4	60.1
Extended parallel corpus, two references	40.9	50.5

Table 4.4: Translation from DGS to German: Effect of extending the parallel corpus

4.3.1.2 Translation from German to DGS

Stein et al. (2012) also built a translation system for the opposite direction, German to DGS. Performance for this direction was worse, something the researchers attributed to the word order of DGS, which permits some degree of variation. Table 4.5 shows the results for the two translation systems PBT and JANE.

System	BLEU	TER
PBT	15.4	77.0
JANE	16.3	76.1

Table 4.5: Translation from German to DGS: Evaluation results

To summarize, the most relevant findings resulting from the translation experiments described in this section are:

[4] The experiments seem to have been performed on data different from that used for the experiments reported in Stein et al. (2012).

- Getting from a sign language (DGS) to a spoken language (German) required a translation step;
- Including an additional German reference that was a more faithful translation of the DGS gloss text improved translation performance;
- Extending the parallel corpus increased translation results; and
- Translating from DGS to German yielded better results than translating in the opposite direction.

4.3.2 Air travel information

Like Stein et al. (2012), Morrissey (2008) translated from sign language into spoken language (ISL to English) and vice versa (English to ISL). She used the Dublin City University in-house Machine Translation using Examples (MaTrEx) system (Stroppa & Way, 2006). Although the name of the system suggests that it relies on example-based machine translation (EBMT) only,[5] the system makes heavy use of SMT techniques also: In its default configuration, MaTrEx is a hybrid SMT/EBMT system, but it is also possible to use either the SMT or the EBMT engine alone. Morrissey (2008) employed only the SMT component. As her data, she used the ATIS Corpus consisting of 595 sentences (cf. Chapter 3). A 90/10 split into training and test data was applied.

4.3.2.1 Translation from sign language to spoken language

For translation from ISL to English, the baseline involved using a lowercase version of the ISL gloss text (i.e., the source side of the test set) as the English hypothesis. This is similar to the sanity check of Stein et al. (2012) (cf. Section 4.3.1) and to the baseline I used for my own experiments in sign language machine translation (cf. Section 4.4). For Morrissey (2008), the difference between using a lowercase version of the source side of the test set and using the (English) output of the MaTrEx system as hypothesis amounted to 26.43 BLEU (25.20 vs. 51.63), as shown in Table 4.6 ("baseline" vs. "MaTrEx (SMT only)").

[5] EBMT denotes a third paradigm of machine translation, which has become less important in recent years.

System	BLEU	WER	PER
Baseline	25.20	60.31	50.42
MaTrEx (SMT only)	51.63	39.32	29.79
MaTrEx (SMT only with distortion limit 10)	52.18	38.48	29.67

Table 4.6: Translation from ISL to English: Evaluation scores

In MaTrEx, a distortion limit can be set, i.e., the maximum number of jumps (block movements) permitted when recombining the target language output can be adjusted. The default is zero jumps. Morrissey (2008) found that a distortion limit of ten jumps worked best when translating from ISL to English. Table 4.6 shows how using this distortion limit aided performance: The BLEU score increased from 51.63 to 52.18.

Morrissey (2008) also carried out experiments on the spoken language part of the ATIS Corpus[6] and found that the results for the translation direction German to English were on a similar order as those for ISL to English, as shown in Table 4.7: The BLEU score was 52.18 for the ISL-to-English system (also shown in Table 4.6) and 60.73 for the German-to-English system. Morrissey (2008) concluded from this that "data-driven MT [machine translation, S.E.] for sign languages […] can achieve automatic evaluation scores comparable to mainstream spoken language MT" (p. 122).

System	BLEU	WER	PER
ISL to English	52.18	38.48	29.67
German to English	60.73	26.59	22.16

Table 4.7: Translation ATIS Corpus: Evaluation scores

Morrissey (2008) extended her experiments to include translation between other sign language/spoken language pairs. The results for her experiments in translating from ISL to German, DGS to English, and DGS to German are displayed in Table 4.8. They were obtained again using only the SMT component of MaTrEx and a distortion limit of 10. For comparison, the table includes the results obtained with this configuration for translation from ISL to English as reported in Tables 4.6 and 4.7.

[6]The original ATIS Corpus (Hemphill, Godfrey, & Doddington, 1990) contained only spoken languages. The extended ATIS Corpus (Bungeroth et al., 2008) also contains ISL, DGS, and South African Sign Language.

System	BLEU	WER	PER
ISL → English:			
MaTrEx	52.18	38.48	29.67
RWTH	**52.62**	**37.63**	**28.34**
ISL → German:			
MaTrEx	39.69	47.25	**38.47**
RWTH	**40.40**	**46.40**	38.58
DGS → English:			
MaTrEx	**48.40**	**41.37**	**30.88**
RWTH	43.16	46.32	31.36
DGS → German:			
MaTrEx	**42.09**	50.31	39.53
RWTH	35.69	**49.15**	**38.68**

Table 4.8: Translation between different sign language/spoken language pairs: Evaluation scores

The table also shows the results for experiments performed on the same data with one of the RWTH Aachen systems described in Section 4.3.1. For each language pair and evaluation metric, the better of the two scores is printed in bold. The table shows that the RWTH system outperformed MaTrEx for translation from ISL to English with regard to all three evaluation metrics (BLEU, WER, and PER) as well as for translation from ISL to German and from DGS to German with regard to two out of three metrics. MaTrEx outperformed the RWTH system for translation from DGS to English with regard to all three metrics.

4.3.2.2 Translation from spoken language to sign language

Morrissey (2008) also performed experiments in translation from spoken language to sign language: from English to ISL, from German to ISL, from English to DGS, and from German to DGS. The results are given in Table 4.9, again obtained with the SMT component of MaTrEx. Similar to the findings of Stein et al. (2012) for translation between German and DGS, Morrissey (2008) observed lower scores when translating from English to ISL than from ISL to English (38.85 vs. 52.18 BLEU).

To summarize, the findings presented in this section are:

- Sign language machine translation produced results on the order of those achieved for spoken language machine translation;

System	BLEU	WER	PER
English → ISL	38.85	46.02	34.33
German → ISL	25.65	57.95	46.62
English → DGS	49.77	45.09	29.59
German → DGS	47.29	45.90	28.67

Table 4.9: Translation from spoken language to sign language: Evaluation scores

- Translation from ISL to English yielded higher scores than translation from English to ISL.

4.4 Automatically translating German train announcements into DSGS

The goal of the project associated with the thesis at hand was to automatically translate written German train announcements of the SBB into synthesized DSGS. The SBB train announcements are parametrized in that they are based on templates with slots, where slots are, e.g., the names of train stations, types of trains, or reasons for delays. Examples 4.1 to 4.3 show templates underlying the German train announcements along with sample instantiations.

Example 4.1. Gleis [Gleisnr.]: Einfahrt des/der [Zugtyp] nach [Ziel], Abfahrt [Uhrzeit]
('Platform [platform no.]: arrival of the [type of train] to [destination], departure [departure time]')
Gleis 10: Einfahrt des RegioExpress nach Burgdorf, Herzogenbuchsee, Langenthal, Olten, Abfahrt 9 Uhr 07
('Platform 10: arrival of the RegioExpress to Burgdorf, Herzogenbuchsee, Langenthal, Olten, departure 9.07 a.m.')

Example 4.2. Der/die [Zugtyp] nach [Ziel], Abfahrt um [Uhrzeit], fällt aus.
('The [type of train] to [destination], departure at [departure time], has been cancelled.')
Die S 22 nach Neuhausen, Schaffhausen, Thayngen, Abfahrt um 23 Uhr 53, fällt aus.
('The S 22 to Neuhausen, Schaffhausen, Thayngen, departure at 11.53 p.m., has been cancelled.')

Example 4.3. Der hintere Zugsteil in den Sektoren [Sektorname] und [Sektorname] verkehrt nur bis [Ziel].
('The back part of the train in Sectors [sector name] and [sector name] is running only as far as

[destination].')
Der hintere Zugsteil in den Sektoren C und D verkehrt nur bis Zürich Hauptbahnhof.
('The back part of the train in Sectors C and D is running only as far as Zurich main station.')

When automatically translating announcements of this kind, one possibility is to take into account precisely their parametrized nature. This was the approach chosen by Segouat (2010), who built a system that converts French train announcements to French Sign Language (*Langue des Signes Française*) (LSF) animations and displays them on a monitor in a train station. The system relies on parallel data consisting of written French announcements on the source side and LSF animations on the target side, both as templates with slots. At runtime, the system identifies the template underlying the input segment and searches for the corresponding LSF animation template. Subsequently, it fills the slots on the target side with the help of further written French/LSF animation correspondences. A coarticulation model is applied to ensure smooth transitions between surrounding and embedded animations.

This approach works well for train announcements. However, when dealing with domains in which content is of non-parametrized nature, a formal understanding of the structure of a sign language is paramount. While a considerable amount of linguistic research has been carried out for DSGS (cf. Chapter 2), there is no reference grammar for this language, i.e., no comprehensive description of the linguistic structure that could serve as a basis for deriving linguistically motivated rules. Since my goal was to build a translation system that can later be extended to other domains, I applied a translation paradigm that does not rely on explicit linguistic knowledge: SMT. Clearly, when extending an SMT system to a broader domain, a considerable amount of data is needed to train the system. Here, possible support comes in the form of sign language recognition, which can help to speed up the process of creating sign language data by means of (semi-)automatic annotation (Dreuw & Ney, 2008).

As data for the current SMT system, I used the parallel corpus of train announcements described in Section 3.6. The corpus was divided into a training set, a development set, and a test set. As in the SMT experiments reported in Sections 4.3.1 and 4.3.2, I used glosses as representation of the sign language side in the machine translation system. In addition, I introduced an approach that automatically generates non-manual information from a string of glosses. This approach relies on sequence classification and is described in Section 4.5. Figure 4.1 visualizes the overall pipeline that transforms a written German train announcement into a DSGS animation: The machine translation system receives as input a German announcement such as *Ausfallmeldung zur*

S1 nach Luzern (‘Notice of cancellation regarding the S1 to Lucerne’), which it translates into DSGS glosses: MELDUNG IX BAHN S1 NACH LUZERN AUSFALL (‘NOTICE IX TRAIN S1 TO LUCERNE CANCELLATION’). The glosses in turn serve as input for the sequence classification system, which produces information pertaining to two non-manual components, eyebrows and head. The output of the machine translation and the sequence classification system is then combined and converted into motion data for the avatar. The animation process is described in more detail in Chapter 5.

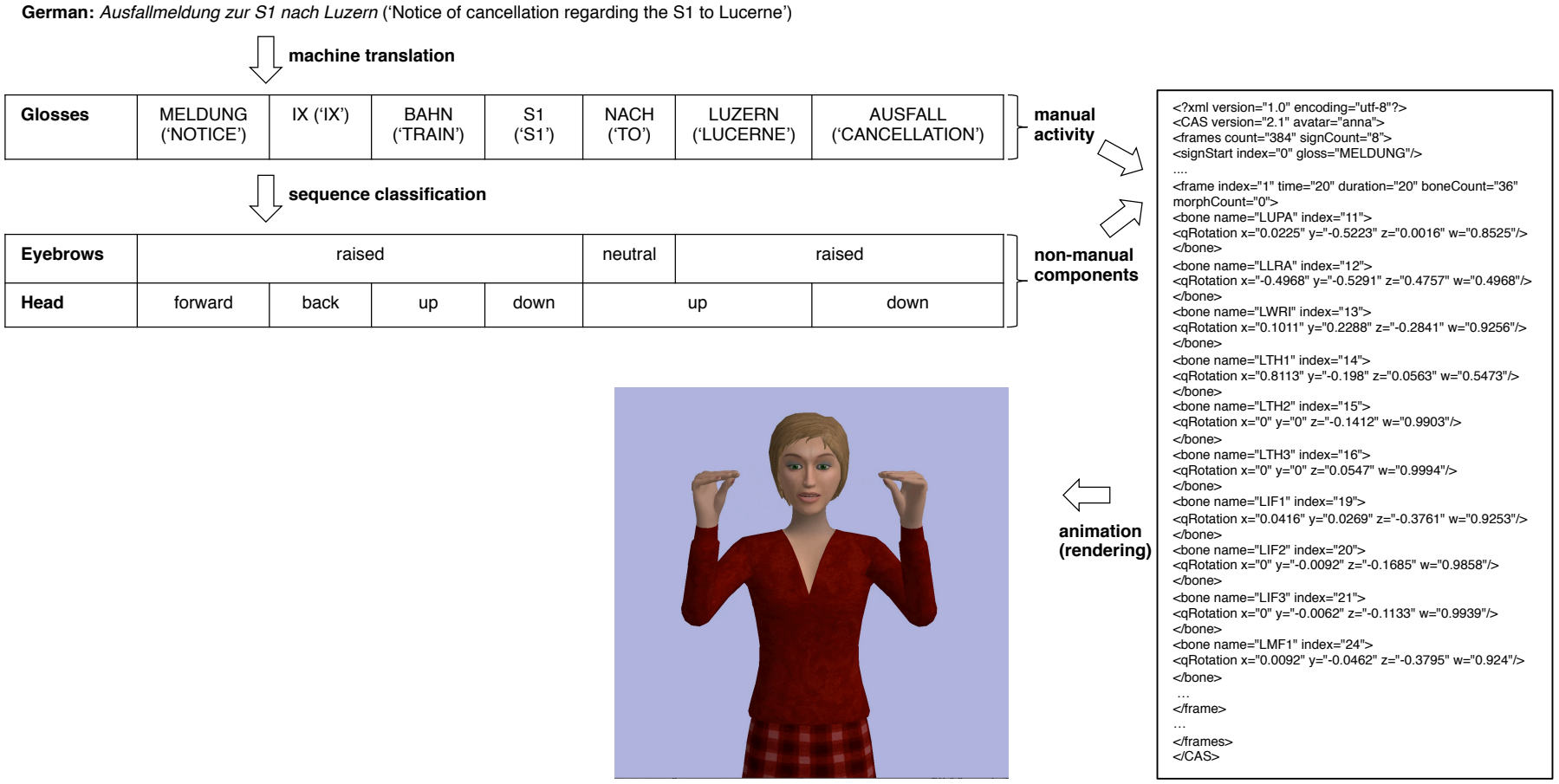


Figure 4.1: Sign language processing pipeline: Machine translation, sequence classification, and animation

As a preprocessing step to the machine translation experiments, I combined multi-word units on the German side of the parallel corpus into single words so that they aligned better with the DSGS side, where these units were commonly represented through one sign token. For example, *S 32* was turned into *S32*, *Interlaken Ost* into *Interlaken-Ost*, and *Zürich Hauptbahnhof* into *Zürich-Hauptbahnhof* to match the DSGS glosses S-32, INTERLAKEN-OST, and ZÜRICH-HAUPTBAHNHOF. I then tokenized and truecased the German side of the training, development, and test sets, leaving the DSGS side untouched. Additionally, a search for overly long (threshold: 80 tokens) or misaligned sentences in the training set was carried out. No sentence pair was removed as a result of this.

I then trained an SMT system with Moses (Koehn et al., 2007). As a language modelling toolkit, I used IRST LM (Federico & Cettolo, 2007). The data for training the language model consisted of the target side of the training set. The n-gram order of the language model was 3. Improved Kneser-Ney smoothing ("improved-shift-beta") (Chen & Goodman, 1996) was applied. For word alignment, the heuristic "grow-diag-final-and" was chosen. These settings correspond to those of the default Moses system.[7]

My baseline consisted of using a lowercased version of the source side of the test set (German text) as hypothesis instead of the DSGS translation output, similar to what had been done in the experiments described in Sections 4.3.1 and 4.3.2. Hence, the baseline configuration was as follows:

- **Input:** German (source side of test set)
- **Hypothesis:** German (source side of test set), lowercased
- **Reference:** DSGS (target side of test set), lowercased

The results of the experimental and the baseline configuration are shown in Table 4.10: The BLEU score was 90.07 for the experimental approach. A NIST score of 9.33 was obtained; WER was 5.13 and PER 4.15. The BLEU score for the baseline approach was 0. Recall from Equation 4.2 in Section 4.2 that this occurs if one of the factors in the overall computation of the score is 0. The NIST score for the baseline setting was 0.17; the corresponding WER and PER scores were 74.37 and 73.95.

[7] `http://www.statmt.org/moses/?n=moses.baseline` (last accessed December 3, 2015).

System	BLEU	NIST	WER	PER
Baseline	0.00	0.17	74.37	73.95
Experimental approach	90.07	9.33	5.13	4.15

Table 4.10: Machine translation of train announcements: Evaluation scores

The exceptionally high performance of my translation system is due to the fact that, as is the case for the systems described in Sections 4.3.1 and 4.3.2, my system operates on a limited domain. Compared to weather reports and air travel information, train announcements are even more restricted with regard to their syntax and their vocabulary. The results reported in this section provide further evidence that statistical sign language machine translation can work well on limited domains, despite the lack of availability of large amounts of training data.

An inspection of the announcements translated with the experimental system showed that candidate and reference translations were identical for 204 of the 300 announcements in the test set (i.e., 68% of the announcements). A selection of these perfect translations is shown in Examples 4.4 to 4.8 along with the German input announcements.[8]

Example 4.4. Gleis 3: InterCity nach Sargans, Landquart, Chur, Abfahrt 21 Uhr 37
('Platform 3: InterCity to Sargans, Landquart, Chur, departure 9:37 p.m.')
GLEIS 3 IX BAHN IC NACH SARGANS LANDQUART CHUR IX ABFAHRT 21 UHR 37
('PLATFORM 3 IX TRAIN IC TO SARGANS LANDQUART CHUR IX DEPARTURE 21 CLOCK 37')

Example 4.5. Information zur S3 nach Wetzikon
('Information regarding the S3 to Wetzikon')
INFO IX BAHN S3 NACH WETZIKON
('INFORMATION IX TRAIN S3 TO WETZIKON')

Example 4.6. Die S5 nach Hardbrücke, Oerlikon, Rafz, Abfahrt 20 Uhr 37, wird verkürzt geführt.
('The S5 to Hardbrücke, Oerlikon, Rafz, departure 8:37 p.m., is operating in shortened form.')

[8] Information appended to the gloss, e.g., regarding the precise nature of an indexical (pointing) sign, has been removed to increase readability.

IX BAHN S5 NACH HARDBRÜCKE OERLIKON RAFZ IX ABFAHRT 20 UHR 37 IX BAHNWAGEN DANN VERKÜRZEN
('IX TRAIN S5 TO HARDBRÜCKE OERLIKON RAFZ IX DEPARTURE 20 CLOCK 37 IX TRAIN-WAGGON THEN SHORTEN')

Example 4.7. Ausfallmeldung zur S12 nach Brugg
('Notice of cancellation regarding the S12 to Brugg')
MELDUNG IX BAHN S12 NACH BRUGG AUSFALL
('NOTIFICATION IX TRAIN S12 TO BRUGG CANCELLATION')

Example 4.8. Nächste Einfahrt: RegioExpress nach Escholzmatt, Schüpfheim, Abfahrt 22 Uhr 12
('Next arrival: RegioExpress to Escholzmatt, Schüpfheim, departure 10.12 p.m.')
NÄCHSTE EINFAHRT IX REGIO-EXPRESS NACH ESCHOLZMATT SCHÜPFHEIM IX ABFAHRT 22 UHR 12
('NEXT ARRIVAL IX REGIO-EXPRESS TO ESCHOLZMATT SCHÜPFHEIM IX DEPARTURE 22 CLOCK 12')

During decoding, the system encountered 54 unknown words. Among them were 15 place names. If a translation for a word or phrase cannot be found, the Moses system inserts the source segment. For the unknown place names, this ultimately led to a correct translation: For example, the DSGS translation for the word *Zäziwil* (a place name) was unknown, as a result of which the system inserted the German source word. Since evaluation was performed on a lowercase version of the candidate translation, the difference between *Zäziwil* (German) and ZÄZIWIL (DSGS) was neutralized and the candidate translation string *zäziwil* evaluated as being correct.

4.5 Automatically generating non-manual information

[This section is an extension of Ebling & Huenerfauth (2015).]

As shown in Chapter 2, non-manual components in sign languages are capable of assuming functions at various linguistic levels, thereby constituting an important part of signing. Table 4.11 lists previous data-driven (mostly statistical) approaches to sign language machine translation along with the sign language representations and the non-manual information used. It shows that

Translation direction	**Sign language representation**	**Non-manual**	**Reference**
German → DGS, DGS → German	glosses	–	Stein, Forster, Zelle, Dreuw, & Ney (2010), Stein, Schmidt, & Ney (2012)
English → ISL, German → DGS, English → DGS German → ISL	glosses	–	Morrissey (2008)
Chinese → Taiwanese SL	glosses	–	H.-Y. Su & Wu (2009)
Spanish → Spanish SL	glosses	–	San-Segundo, Lopez, Martin, Sanchez, & Garcia (2010)
Catalan → Catalan SL	glosses	mouth morphemes	Massó & Badia (2010)

Table 4.11: Overview of data-driven approaches to sign language machine translation

sign language was represented almost exclusively with glosses in these approaches. As discussed in Chapter 2, glosses primarily encode information about the manual activities of signing.

It follows from this that non-manual information has not been included in most previous data-driven sign language machine translation systems. Morrissey (2008) acknowledged that "omitting NMFs [non-manual features, S.E.] means some important grammatical and semantic information is absent from the annotations and will thus be absent from translations, ultimately reducing the translation quality" (p. 95). Stein et al. (2012) found non-manual information to be missing upon inspection of their translation results.

An approach that incorporated non-manual information is that of Massó & Badia (2010), which treated "mouth morphemes" ("mouth gestures" in the terminology of Section 2.1) as factors in an SMT system when translating from Catalan into Catalan Sign Language. Although this is not a technical requirement, factors in the factored SMT framework typically represent generalizations over word forms that the system may fall back to in case of unknown words. Common examples of factors are lemmas. Since mouth morphemes represent no such generalizations, including them as factors is not an obvious approach.

As outlined in Chapter 1, few sign language processing applications exist that make use of more than one sign language technology (of which examples are sign language recognition, sign language machine translation, or sign language animation). In particular, statistical sign language machine translation and sign language animation have rarely been combined in the past. Because of this, the fact that previous statistical sign language machine translation systems used a sign language representation that falls short of capturing non-manual information did not pose a problem. However, once the output of a machine translation system is used as input for a sign

language animation system, this lack of richness becomes apparent: Absence of non-manual information in sign language animations has been shown to lead to lower comprehension scores and lower subjective ratings of the animations by Deaf informants (Kacorri, Lu, & Huenerfauth, 2013).

My goal was to include non-manual information in the overall process of translating written German train announcements to synthesized DSGS. More precisely, my aim was to bridge the gap between the output of a sign language translation system and the input of a sign language animation system by including non-manual information in the output of the translation system. Table 4.12 shows a train announcement that includes information on head and eyebrow movement.

Gloss **Gloss: translation**	MELDUNG ('NOTICE')	IX ('IX')	BAHN ('TRAIN')	S1 ('S1')	NACH ('TO')	LUZERN ('LUCERNE')	AUSFALL ('CANCELLATION')
Eyebrows	raised				neutral	raised	
Head	forward	back	up	down	up		down

Table 4.12: DSGS translation of the German train announcement *Ausfallmeldung zur S1 nach Luzern* ('Notice of cancellation regarding the S1 to Lucerne')

One way of considering non-manual information in a translation task is to simply append it to glosses. This representation is shown in Example 4.9 for the announcement introduced in Table 4.12. The non-manual features are printed in bold.

Example 4.9. Ausfallmeldung zur S1 nach Luzern ('Notice of cancellation regarding the S1 to Lucerne'):

MELDUNG__**Head_forward__Eyebrows_raised**

IX__**Head_back__Eyebrows_raised**

BAHN__**Head_up__Eyebrows_raised**

S1__**Head_down__Eyebrows_raised**

NACH__**Head_up__Eyebrows_neutral**

LUZERN__**Head_up__Eyebrows_raised**

AUSFALL__**Head_down__Eyebrows_raised**

However, such a representation aggravates the issue of data sparseness, since the size of the vocabulary is no longer equivalent to the number of unique glosses but to the number of unique combinations of glosses and non-manual features. This increases the likelihood that tokens appear in the decoding phase that have not been seen during training (*out-of-vocabulary items*,

OOV). Such a representation also does not accommodate the multi-level nature of sign languages: Three tiers (glosses, head, and eyebrow information) are collapsed into one.

I propose an approach that schedules the automatic generation of non-manual information after the machine translation step and views it as a sequence classification task. More precisely, the process of generating non-manual information is conceived as the task of labelling glosses (as representations of the manual components) with non-manual features. This conception is justified by the fact that the boundaries of the non-manual components in the DSGS train announcements align with those of manual components, as can be seen in the example in Table 4.12. This is because the non-manual components in our train announcements fulfill linguistic rather than purely affective functions (cf. Section 2.1).

Sequence classification has been used to solve various natural language processing problems, such as part-of-speech tagging and chunking. In contrast to standard classifiers, sequence classifiers are capable of taking into account the sequential nature of data. Sequential Conditional Random Fields (CRFs) (C. Sutton & McCallum, 2012) are a state-of-the-art approach for this. Given one or more sequences of tokens (the *evidence*), CRFs compute the probability of a sequence of labels (the *outcome*). While multiple evidence layers are permitted, CRFs only allow for the prediction of one outcome layer.

The Wapiti toolkit (Lavergne, Cappé, & Yvon, 2010) provides an efficient implementation of CRFs.[9] Sequence classification with Wapiti follows a train–test–evaluate cycle. Hand-crafted *feature templates* are created to specify which tokens of the evidence are considered for the prediction of the outcome labels. In addition, the *emission order* is declared, indicating whether the evidence is conditioned on label unigrams (emission order 1) or bigrams (emission order 2). During the training step, the feature templates are instantiated with the training data.

The parallel corpus of 2 986 German/DSGS train announcements described in Section 3.6 was used to perform the sequence classification experiments in Wapiti. The data had been randomly divided into 10 folds of 300 announcements each to enable 10-fold cross validation. For each validation round, eight folds were used for training, one was used for development, and one for testing. Using the ground truth as opposed to the machine translation output as data was

[9] `http://wapiti.limsi.fr/manual.html`

motivated by an interest in investigating the potential of sequence classification in isolation, without possible error propagation from the preceding machine translation step.

4.5.1 Experiment configurations

The aim of the experiments described here was to predict the most probable sequence of non-manual features for a sequence of glosses. As previously stated, CRFs allow for the prediction of one outcome layer at a time. Hence, the two label layers head and eyebrows could either be collapsed into a single label (Configuration G→H+E, Table 4.13), or a separate classifier could be trained for each feature (Configurations G→H and G→E, Table 4.14). A downside of Configuration G→H+E is that there is a potential for data sparseness, as the number of possible outcome labels is equivalent to the number of cross-combinations of head and eyebrow labels occurring in the training data. However, even with this approach, the risk of data sparseness is lower than that of appending the non-manual features to the sign language glosses during the machine translation task, as previously discussed.

Evidence **Gloss**	**Label** **Non-manual**
MELDUNG (‘NOTICE’)	forward_raised
IX (‘IX’)	back_raised
BAHN (‘TRAIN’)	up_raised
S1 (‘S1’)	down_raised
NACH (‘TO’)	up_neutral
LUZERN (‘LUCERNE’)	up_raised
AUSFALL (‘CANCELLATION’)	down_raised

Table 4.13: Configuration G→H+E: Collapsing the two label layers head and eyebrows into a single label

With Configurations G→H and G→E, each label layer (head and eyebrows, respectively) is treated in isolation, which means that dependencies between the two are not captured. However, conceptually, dependencies between the two types of non-manual information exist in that they assume specific linguistic functions together, e.g., topicalization, rhetorical questions, or conditional expressions in DSGS (cf. Section 2.1). These dependencies can be accounted for by introducing a cascaded approach, i.e., by using the output of one classifier as additional input for the other. More precisely, the output of the head classifier can be used as additional evidence for the eyebrow classifier and vice versa. This is shown as Configurations G_E→H and G_H→E in Table 4.15. Note that such a representation accommodates the multi-level nature of sign languages.

Evidence **Gloss**	**Label** **Head**
MELDUNG (‘NOTICE’)	forward
IX (‘IX’)	back
BAHN (‘TRAIN’)	up
S1 (‘S1’)	down
NACH (‘TO’)	up
LUZERN (‘LUCERNE’)	up
AUSFALL (‘CANCELLATION’)	down

Evidence **Gloss**	**Label** **Eyebrows**
MELDUNG (‘NOTICE’)	raised
IX (‘IX’)	raised
BAHN (‘TRAIN’)	raised
S1 (‘S1’)	raised
NACH (‘TO’)	neutral
LUZERN (‘LUCERNE’)	raised
AUSFALL (‘CANCELLATION’)	raised

Table 4.14: Configurations G→H (top) and G→E (bottom): Training a separate classifier for each of the two features head (top) and eyebrows (bottom)

To better model the sequential dependencies in a given data set, an IOB representation (Sang & Veenstra, 1999) can be used. In this format, B denotes the first token of a label sequence, I a sequence-internal token, and O is used for tokens that are not part of a sequence of a label under consideration. This format is applied as Configurations G→H_{IOB} and G→E_{IOB} (Table 4.16). Note that in the case at hand, O does not occur, since the data contains multi-class as opposed to binary annotations and *neutral* is one of the possible class labels.

For my experiments, I applied all of the above seven configurations, as summarized in Table 4.17.

Among the strengths of CRFs is their ability to handle a large amount of features and cope with redundancy (Lavergne et al., 2010). 26 feature templates similar to the templates used by Roth & Clematide (2014) were provided for each evidence layer. The overall context ranged from the three previous tokens to the three following tokens relative to the current position. Each window was included with both emission order 1 (unigram) and 2 (bigram). In addition, raw unigram and bigram output distribution were included.

Evidence		Label
Gloss	**Eyebrows**	**Head**
MELDUNG (‘NOTICE’)	raised	forward
IX (‘IX’)	raised	back
BAHN (‘TRAIN’)	raised	up
S1 (‘S1’)	raised	down
NACH (‘TO’)	neutral	up
LUZERN (‘LUCERNE’)	raised	up
AUSFALL (‘CANCELLATION’)	raised	down

Evidence		Label
Gloss	**Head**	**Eyebrows**
MELDUNG (‘NOTICE’)	forward	raised
IX (‘IX’)	back	raised
BAHN (‘TRAIN’)	up	raised
S1 (‘S1’)	down	raised
NACH (‘TO’)	up	neutral
LUZERN (‘LUCERNE’)	up	raised
AUSFALL (‘CANCELLATION’)	down	raised

Table 4.15: Configurations G_E→H (top) and G_H→E (bottom): Using the output of the eyebrow classifier as additional evidence for the head classifier (top) and vice versa (bottom)

4.5.2 Results

Table 4.18 shows the results of the experiments obtained using the default settings of Wapiti. “Experimental approach” refers to the configurations described in Section 4.5.1. The lower baseline for each configuration consisted of applying a (non-sequential) Maximum Entropy classifier also offered in Wapiti. This implied regarding each token as a sequence of its own. Hence, with this classifier, token error and sequence error are identical.

For each experimental or baseline approach, Table 4.18 provides the following numerical information:

- Number of labels
- Token error: This is the mean of the token errors of the ten rounds of a 10-fold cross validation. The token error for an individual validation round is calculated as the percentage of incorrectly predicted labels.
- Standard deviation of token error for the ten rounds
- Confidence interval of token error: This is the confidence interval at a confidence level of 95% calculated over the mean of the token errors using Student’s t-test.

Evidence **Gloss**	**Label** **Head**
MELDUNG (‘NOTICE’)	B_forward
IX (‘IX’)	B_back
BAHN (‘TRAIN’)	B_up
S1 (‘S1’)	B_down
NACH (‘TO’)	B_up
LUZERN (‘LUCERNE’)	I_up
AUSFALL (‘CANCELLATION’)	B_down

Evidence **Gloss**	**Label** **Eyebrows**
MELDUNG (‘NOTICE’)	B_raised
IX (‘IX’)	I_raised
BAHN (‘TRAIN’)	I_raised
S1 (‘S1’)	I_raised
NACH (‘TO’)	B_neutral
LUZERN (‘LUCERNE’)	B_raised
AUSFALL (‘CANCELLATION’)	I_raised

Table 4.16: Configurations G→H_{IOB} (top) and G→E_{IOB} (bottom): Applying an IOB format

Configuration	**Evidence**	**Label**
G→H+E	glosses	head and eyebrows
G→H	glosses	head
G→E	glosses	eyebrows
G_E→H	– glosses – eyebrows	head
G_H→E	– glosses – head	eyebrows
G→H_{IOB}	glosses	head IOB
G→E_{IOB}	glosses	eyebrows IOB

Table 4.17: Overview of configurations

- Sequence error: This is the mean of the sequence errors of a 10-fold cross validation. The sequence error for an individual validation round is calculated as the percentage of incorrectly predicted sequences, i.e., sequences containing at least one token error.
- Standard deviation of sequence error
- Confidence interval of sequence error: This is the confidence interval at a confidence level of 95% calculated over the mean of the sequence errors using Student’s t-test.

Configuration	**Labels**	**Token level**			**Sequence level**		
		Token error (%)	Standard dev.	Conf. interval	Sequence error (%)	Standard dev.	Conf. interval
Predicting H+E	**31**						
G→H+E		1.88	0.50	0.36	10.43	2.53	1.81
— Lower baseline		14.99	0.53	0.38	14.99	0.53	0.38
Predicting H	**13**						
G→H		1.62	0.45	0.32	8.96	2.43	1.7
— Lower baseline		12.90	0.53	0.38	12.90	0.53	0.38
G_E→H		1.62	0.50	0.36	9.19	2.40	1.72
— Upper bound		1.29	0.39	0.28	7.86	2.05	1.46
— Lower baseline		12.96	0.53	0.38	12.96	0.53	0.38
Predicting E	**3**						
G→E		0.74	0.24	0.17	6.85	1.81	1.29
— Lower baseline		4.98	0.45	0.32	4.98	0.45	0.32
G_H→E		0.66	0.16	0.12	5.72	0.96	0.69
— Upper bound		0.45	0.11	0.08	4.21	0.88	0.63
— Lower baseline		5.03	0.46	0.33	5.03	0.46	0.33
Predicting H_{IOB}	**21**						
G→H_{IOB}		1.81	0.56	0.40	9.13	2.84	2.03
— Lower baseline		19.40	0.75	0.54	19.40	0.75	0.54
Predicting E_{IOB}	**6**						
G→E_{IOB}		1.41	0.30	0.21	9.96	1.85	1.33
— Lower baseline		18.54	0.70	0.50	18.54	0.70	0.50

Table 4.18: Sequence classification experiments: Results

The results in Table 4.18 show that the experimental approaches achieved a lower sequence error rate than the baselines (with the difference being greater than the confidence interval of the values) in all but two cases. The two exceptions for which the sequence error rate of the experimental approach was higher than that of the baseline approach were G→E (experimental: 6.85%; baseline: 4.98%) and G_H→E (experimental: 5.72%; baseline: 5.03%). For these cases, applying a sequential rather than a standard (non-sequential) classifier did not aid performance. The error rates of the experimental approaches are notably low, which is at least partly due to the nature of the data used for the experiments: As described in Section 4.4, the SBB train announcements are highly parametrized in that they are based on a limited set of phrasal templates. The comparison of experimental approaches and baselines is visualized in Figure 4.2. In what follows, the results are discussed in more detail.

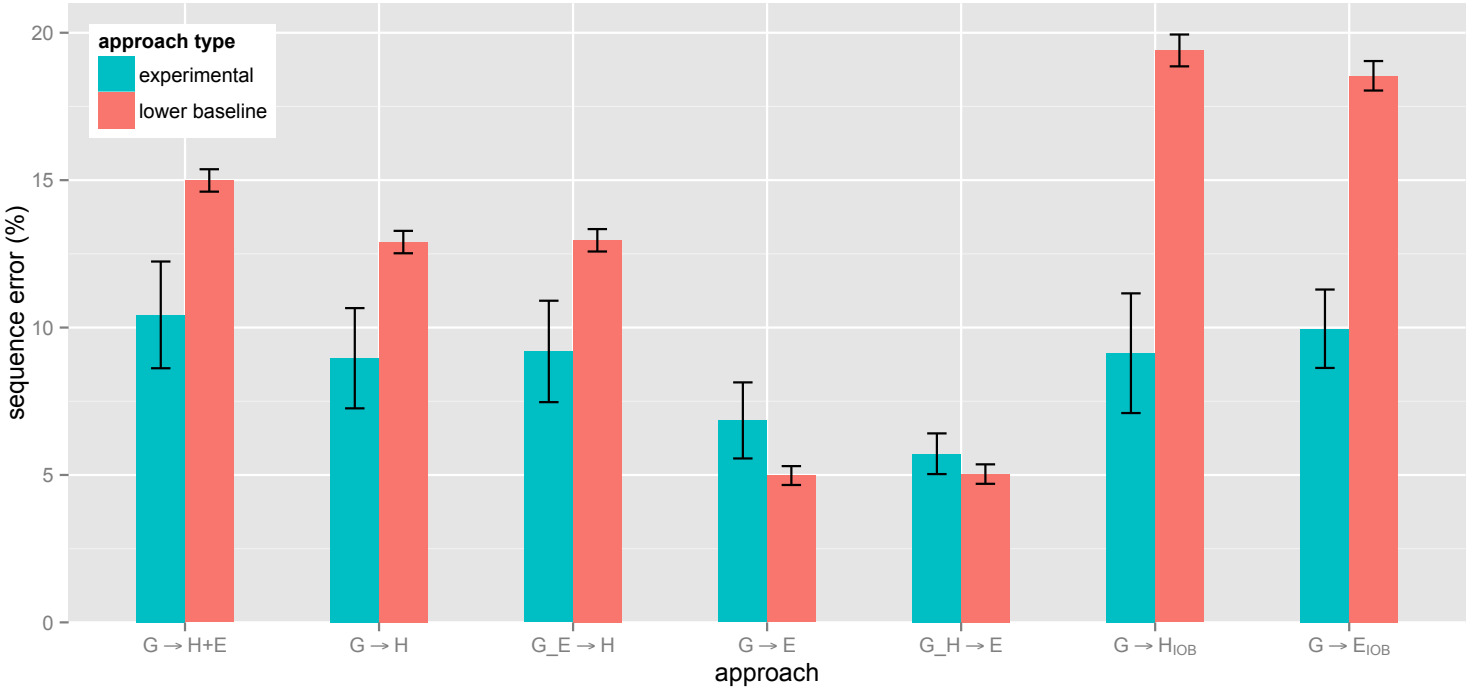


Figure 4.2: Comparison of sequence error rates of experimental and lower baseline approaches

4.5.2.1 Cascaded vs. non-cascaded

Between Configuration G→H (non-cascaded) and G_E→H (cascaded), both predicting head information, Configuration G→H exhibited a lower sequence error rate (8.96% vs. 9.19%). Between Configuration G→E (non-cascaded) and G_H→E (cascaded), both predicting eyebrow information, Configuration G_H→E achieved a lower sequence error rate (5.72% vs. 6.85%).

To examine the theoretical potential of the cascaded approach, I determined the upper bound, i.e., the result of applying the model learned from the training data on the ground-truth data. In other words, as data for the additional evidence layer (eyebrow information for Configuration G_E→H and head information for Configuration G_H→E), the gold-standard annotations of these layers instead of the output of Configurations G→E and G→H, respectively, were used. The resulting numbers are shown in the table as "Upper bound" for Configurations G_E→H and G_H→E. Configuration G_E→H/Upper bound achieved a lower sequence error rate than Configuration G→H (7.86% vs. 8.96%). Configuration G_H→E/Upper bound also achieved a lower sequence error rate than Configuration G→E (4.21% vs. 6.85%); here, the magnitude of the difference was greater than the confidence intervals of the values. These results show that a cascaded approach is capable of outperforming a non-cascaded approach, and they imply that in DSGS, head information provides more useful information for predicting eyebrow information than vice versa. Figure 4.3 visualizes the comparison of cascaded and non-cascaded approaches.

4.5.2.2 IOB vs. non-IOB

Between Configuration G→H (non-IOB format) and G→H_{IOB} (IOB format), both predicting head information, Configuration G→H produced a lower sequence error rate (8.96% vs. 9.13%). Between Configuration G→E (non-IOB format) and G→E_{IOB} (IOB format), both predicting eyebrow information, Configuration G→E yielded a lower sequence error rate (6.85% vs. 9.96%). In this case, the magnitude of the difference was greater than the confidence interval of the values. These results show that applying an IOB format was not beneficial for the task at hand, most likely due to data sparseness: Introducing the IOB format doubled the number of labels for Configuration G→E_{IOB} compared to Configuration G→E (6 vs. 3 labels, cf. Table 4.18), while the relative increase was smaller for Configuration G→H_{IOB} compared to Configuration G→H (21 vs. 13 labels), indicating that five head features appeared sequence-initially only, i.e., spanned over one gloss. Figure 4.4 visualizes the comparison of IOB and non-IOB approaches.

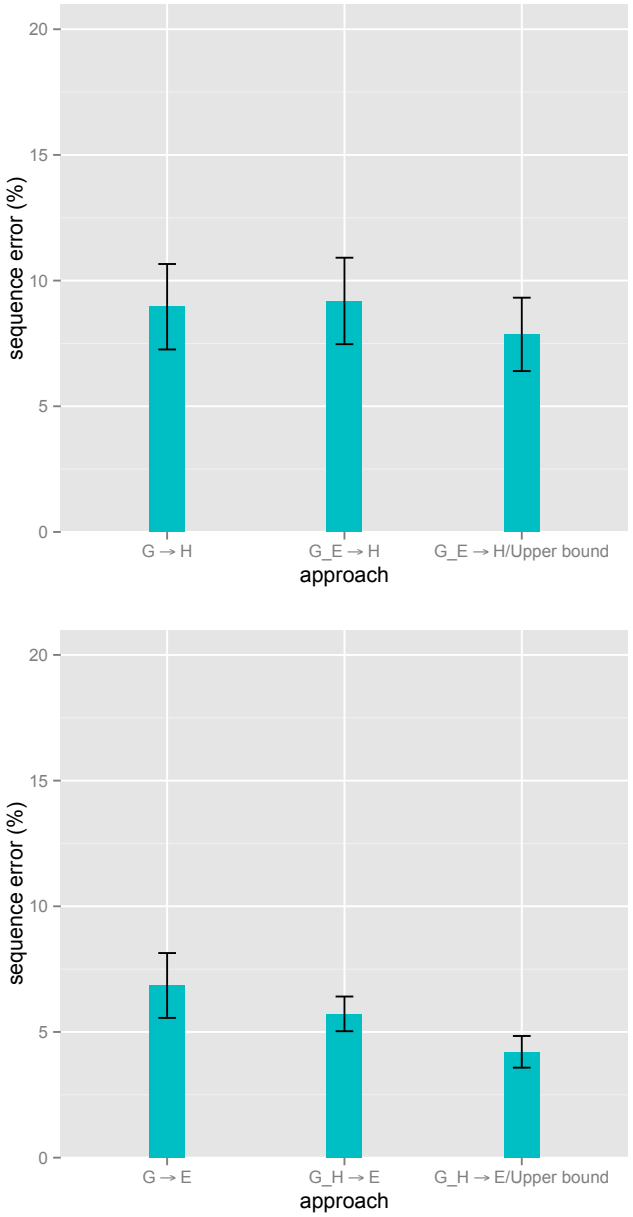


Figure 4.3: Sequence error rates of cascaded vs. non-cascaded approaches: Predicting head information (top) and eyebrow information (bottom)

4.5.2.3 Analysis of features

An examination of the 50 highest-weighted (instantiated) features in the models of the experimental approaches of Configurations G→H+E, G_E→H, and G_H→E for the first round of the 10-fold cross validation showed that among the highest-weighted features for Configuration G→H+E were 31 bigram features and 19 unigram features. The most frequently occurring feature context window consisted of the current token of the (gloss) evidence layer (i.e., relative

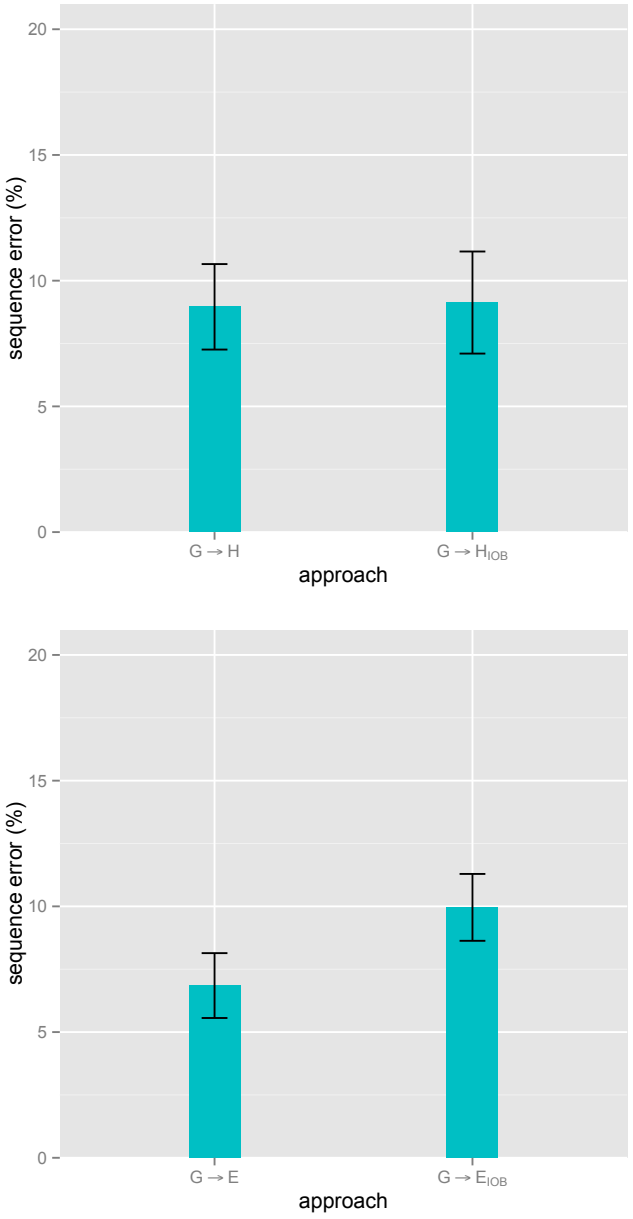


Figure 4.4: Sequence error rates of IOB vs. non-IOB approaches: Predicting head information (top) and eyebrow information (bottom)

position 0). Thus, the identity of a lexical item contributed to the model's prediction of the non-manual feature that co-occurs with it. The second- and third-best performing feature context windows contained the previous token (-1) and the following token (+1) of an evidence layer, respectively. This was followed by a window containing the current and the following token (0 to +1). Thus, the neighboring lexical items contributed to the prediction of the non-manual feature.

For Configuration G_E→H (predicting head information), the 50 top-weighted features consisted of 26 bigram and 24 unigram features. 48 features used tokens from the gloss evidence layer, while 2 used tokens from the added eyebrow information layer. For Configuration G_H→E (predicting eyebrow information), this number was considerably higher: Among the 50 best-scoring features were 27 that used tokens from the head information layer. Again, this serves as evidence that head information is valuable when predicting eyebrow information in DSGS. Here, the most frequently occurring feature context windows included the three previous and the current token (-3 to 0). An instantiation of this pattern is shown in Table 4.19 (in horizontal tier notation rather than in the vertical representation used in Tables 4.13 to 4.16): A sequence of the head moving back/right (-3), up and down (-2), up (-1), and down again (0) is used to predict that the eyebrow label at the current position (0) is "neutral".

Position	-3	-2	-1	0
Gloss **Gloss: translation**	IX ('IX')	ABFAHRT ('DEPARTURE')	8 ('8')	UHR ('CLOCK')
Head	back/right	up and down	up	down

Table 4.19: G_H→E: Feature using the context -3 to 0

4.6 Summary

This chapter has presented my experiments in translating from German to DSGS using the statistical paradigm. An overview of previous statistical sign language machine translation work has been given. In particular, I have discussed two previous approaches that operated on a limited domain (weather reports and air travel information). Domain specificity was precisely one of the reasons why these approaches produced results that were highly satisfactory, especially in light of the comparatively small parallel corpora used to train the systems. My own experiments used data from a restricted domain as well: The train announcements I worked with are highly parametrized, which is why translation on this data worked extraordinarily well.

I have shown that so far, sign language has been represented mainly with glosses in previous statistical sign language machine translation systems. Such a representation is inadequate as it does not capture an important part of signing, namely non-manual information. This inadequacy becomes obvious if the output of a sign language machine translation system is used as input for a sign language animation system, where lack of non-manual information has been shown to be one of the main factors standing in the way of Deaf users' acceptance.

I have presented work that bridges the gap between the output of a sign language machine translation system and the input of a sign language animation system by incorporating non-manual information into the output of the translation system. The approach schedules the generation of non-manual information after the machine translation task and treats it as a sequence classification task. Sequence classification is a technique commonly used in the automatic processing of spoken languages. As far as I can see, my work is the first to apply it to sign languages.

The experimental approaches consisted of predicting head and eyebrow information together in one label, predicting head and eyebrow information separately, predicting head information by using eyebrow information as additional evidence and vice versa (cascaded approach), and applying an IOB format. The experimental approaches outperformed the baselines (non-sequential classifiers) in all but two cases. The results underlined the potential of a cascaded approach, i.e., of using the output of one classifier as additional input for another. In particular, they suggested that for DSGS, head information is more valuable for predicting eyebrow information than vice versa.

As systems translating into sign language increasingly include non-manual information, future work will have to focus on the development of an automatic evaluation metric that takes into account the multi-level nature of sign languages. A more distant goal in statistical sign language machine translation will be to build systems for less restricted domains than the railway, weather, and air travel domain. These systems will have to be capable of dealing with a greater variety of phenomena typical of sign languages. For example, sign languages feature a number of types of signs that are not stable ("frozen") units but for which one or several of the manual parameters (hand shape, hand position, location, and movement) are determined by the context. Among these phenomena are *spatial verbs* and *agreement verbs* (Padden, 1988). Both types of verbs have the hand shape as the only fixed parameter; the parameters hand position, location, and movement can be modified. With spatial verbs, the execution of the latter parameters is determined by the source and/or goal of an action as previously identified by reference to a point or direction in the signing space. For example, the execution of the hand position, location, and movement parameters in the spatial verb GEHEN ('GO') in DSGS depends on *from where to where* a referent is going. With agreement verbs, the execution of these parameters is determined by the subject and/or object of a signed sentence as previously identified by reference to a point or direction in the signing space. For example, the execution of the hand position, location, and movement parameters in the agreement verb GEBEN ('GIVE') in DSGS depends on *who is giving to whom*. Spatial and agreement verbs, along with other context-dependent

phenomena, are likely to appear in sign language data that is of less controlled nature than are train announcements, weather reports, and air travel announcements.[10]

Future work in automatic generation of non-manual information through sequence classification might look into dealing with non-manual information that is not lexically cued, i.e., not recoverable from the glosses alone. For example, recall from Section 2.1 that an interrogative sentence in DSGS can have the same surface form (gloss order) as a declarative sentence.[11] Thus, to disambiguate between the two sentence types, information from the German source sentence could be exploited, e.g., question marks could be included as absolute features. Leveraging information from the source sentence would also make it possible to capture instances in which a grammatical function is expressed non-manually only. For example, in American Sign Language (ASL), it is possible to convey negation solely via head shake, without the use of any manual activity (Neidle, Kegl, MacLaughlin, Bahan, & Lee, 2001).

[10] The difference between spatial verbs and agreement verbs has previously been captured through a distinction between a *topographic* (spatial verbs) and a *syntactic* (agreement verbs) use of the signing space (Poizner, Klima, & Bellugi, 1987), though agreement verbs arguably make use of the topographic signing space as well (Konrad, 2010).

[11] Interrogative sentences did not occur in the train announcement data I worked with.

Chapter 5

Sign language animation

[This chapter is an extension of Ebling (2013), Ebling & Glauert (2013), and Ebling & Glauert (2015).]

Sign language animation, the process of creating a signing avatar, is a young field of research, looking back on about 20 years of existence (Kipp, Heloir, & Nguyen, 2011). In contrast to videos of human signers, sign language animations are capable of providing an anonymous representation of a signer. This minimizes the likelihood of legal implications arising from, e.g., display on the web. Moreover, the content of a sign language animation can typically be modified more easily than that of a self-contained video. Using sign language animation also bears the possibility of tailoring the avatar's appearance (gender, level of formality from serious to cartoon-like, etc.) and speed of signing to a user's needs. In addition, it is often possible for the user to directly adjust the point of view of the avatar as shown in Figure 5.1. Sign language animations also require lower bandwidth than videos (Glauert, 2013).

Figure 5.1: Points of view

5.1 Approaches

Sign language animations may be created through three different approaches: hand-crafted animation, motion capturing, or synthesis from form notation (Glauert, 2013). JASigning, the avatar system I used to synthesize Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS) train announcements (cf. Section 5.3), relies on synthesis from form notation. In what follows, each of the three approaches is described in turn. Knowledge of the hand-crafted animation and the motion capturing approach is essential for understanding the advantages and disadvantages of the synthesis-from-form-notation approach.

Hand-crafted animation consists of manually modelling and posing an avatar character in an animation software such as Maya, 3ds Max, or Blender. This procedure typically yields good results but is also very labor-intensive. Figure 5.2 shows the hand-crafted signing avatar Pedro created for the 2007 World Federation of the Deaf Congress in Spain.[1] A further example of a hand-crafted signing avatar is Paula (McDonald et al., 2013). Section 5.4 describes the process of synthesizing the DSGS finger alphabet using Paula. Paula is shown in Figure 5.12.

Figure 5.2: Hand-crafted signing avatar Pedro

A signing avatar may also be animated based on information obtained from motion capturing, which involves recording a human's signing. Two types of motion capturing exist: active motion capturing, where the signer wears a body suit, gloves, and possibly other equipment with internal sensors (as shown in Figure 5.3 on the left), and passive motion capturing, where the signer wears external markers that are traced by a single or multiple cameras (as shown in Figure 5.3 on the right) (Wolfe, Cook, McDonald, & Schnepp, 2013).[2] Examples of avatars that are based on

[1] `http://www.youtube.com/watch?v=QiY5LU-II6Q` (last accessed October 1, 2015).

[2] `http://www.visicast.cmp.uea.ac.uk/Images/videos/motion_capture.avi` (last accessed: October 1, 2015),
`http://www.mocaplab.com/news/bbc-2-see-hear-visit-mocaplab/` (last accessed October 1, 2015).

motion-captured data are Tessa (Cox et al., 2002) and Sign3D (Lefebvre-Albaret, Gibet, Turki, Hamon, & Brun, 2013). Sign language animations obtained through motion capturing are typically of good quality. A major drawback of this approach is the long calibration time and the extensive postprocessing required: In the case of passive motion capturing, marker positions often need to be post-corrected, which is time-consuming and may still not result in satisfactory output, as some data may be missing (Wolfe et al., 2013). In addition, motion capturing is an invasive technology to begin with: Signing naturally while wearing bulky tracking equipment poses a major challenge. Moreover, the equipment itself is expensive.

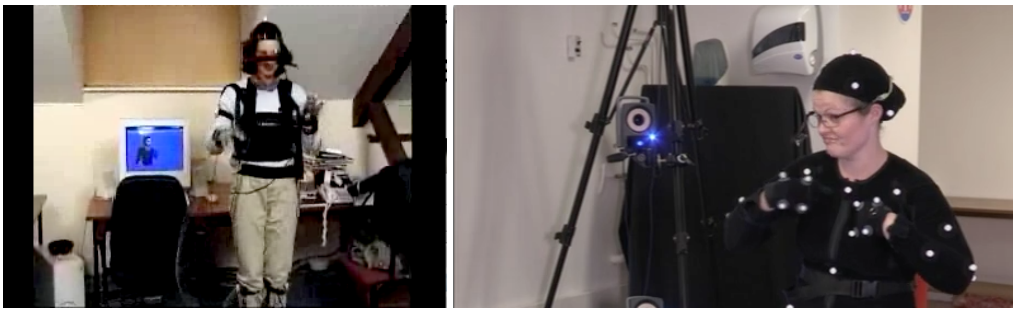


Figure 5.3: Active (left) and passive (right) motion capturing

Both with animation from motion capturing and with hand-crafted animation, the inventory of available signing comprises precisely the sign forms previously created and their combinations (transitions may be generated through interpolation). The sublexical structure of the signs is not accessible at runtime. Hence, sublexical parameters cannot be modified on the fly. This is different for the synthesis-from-form-notation approach: Here, a full-fledged animation system exists that supports synthesis of any sign form that can be described through the associated notation. Animations are created at runtime from the form notations, which means there is access to the sublexical structure of signs. Therefore, sublexical parameters can be modified on the fly when embedding a sign in a new context. For example, the place of articulation of a sign can be adjusted to take account of coarticulation effects. Recall from Section 4.6 that sign languages also feature phenomena for which one or several of the manual parameters are determined by the context in the first place. The fact that notation-based synthesis allows for signs to be modified in context makes it the most flexible of the three approaches.

Signing avatars synthesized from form notation are able to render dynamic content, e.g., display the sign language output of a machine translation system, present the contents of a sign language wiki or an e-learning application, visualize lexicon entries or present public transportation information (Efthimiou et al., 2012; Kipp, Heloir, & Nguyen, 2011). At the same time, this approach

to sign language animation typically results in the lowest quality: Controlling the appearance of all possible sign forms that may be produced from a given notation is virtually impossible. An example of an animation system based on this approach is JASigning, which relies on the Hamburg Notation System for Sign Languages (HamNoSys) as its form notation. I used JASigning to synthesize DSGS train announcements. The JASigning character Anna is shown in Figure 5.4. The system is described in more detail in Section 5.3.1.

Figure 5.4: JASigning character Anna (figures from Ebling & Glauert, 2015)

5.2 Evaluation

My work in synthesizing DSGS train announcements and DSGS fingerspelling sequences as reported in Sections 5.3 and 5.4 included evaluating the quality of the resulting animations. No automatic procedure exists for assessing the quality of signing avatars. Sign language animation evaluation studies so far have been carried out in the form of user studies. Here, a distinction is typically made between two concepts: the degree to which a user understands the content of an animation (*comprehension*) and the degree to which he or she accepts it (*acceptance*) (Huenerfauth, Zhao, Gu, & Allbeck, 2007). It is important to note that there is some overlap between these two concepts. However, distinguishing between them makes sense in light of the method used to assess each concept: Comprehension is typically assessed through objective comprehension tasks, while acceptance is commonly assessed via subjective participant ratings. The study I conducted to evaluate synthesized DSGS train announcements (Sections 5.3.3) was an acceptance study, while the study aimed at assessing the quality of synthesized DSGS fingerspelling sequences (Section 5.4.2) was a comprehension study.

5.2.1 Comprehension

Several studies assessing the comprehension of signing avatars have been carried out, of which four are listed in Table 5.1. In what follows, the most important methodological contributions and qualitative findings of each of these studies are presented.

Study	**Avatar**	**Language/ communication system**	**Participants**	**Stimuli**
Huenerfauth et al. (2007)	ASL classifier predicate generation system	ASL, Signed English (lower baseline)	15 Deaf	20 animations
Kipp, Heloir, & Nguyen (2011)	EMBR	DGS	13 Deaf	11 signed sequences
Lefebvre-Albaret (2011)	JASigning	LSF	6 Deaf; 5 Deaf, 5 hearing	20 isolated signs, 5 full sentences
Smith & Nolan (2015)	JASigning	ISL	15 Deaf	5 story segments

Table 5.1: Sign language animation comprehension studies

Huenerfauth et al. (2007) asked participants to subjectively rate their comprehension of animated sign sequences. The researchers also included an objective comprehension task, as part of which the participants had to pick among several visualizations the one that most closely represented the situation of an animation. The researchers observed a weak correlation between the subjective ratings and the results of the objective comprehension task: "There appears to be a difference between a respondent's *perceived* understanding and her *actual* understanding of an animation." (Huenerfauth et al., 2007, p. 217) To test actual comprehension, the researchers recommended including an objective task in future studies.

Kipp, Heloir, & Nguyen (2011) applied a *delta evaluation*: The comprehension scores for sign language animations were computed relative to the comprehension scores for videos of human signers performing the same signs. As videos of human signers, the original videos as well as overarticulated remakes were used. Both an objective and a subjective measure of comprehension was applied: For the objective measure, the ratio of the number of glosses from the sign sequence that participants repeated when rendering the content to the overall number of glosses in their rendering was computed. The subjective measure consisted of Deaf experts assessing the participants' comprehension. The original videos received comprehension scores of 71% (objective measure) and 61% (subjective measure). Comprehension of the overarticulated video remakes was higher (82% for the objective measure). The animations received scores of 58.4% (objective) and 58.6% (subjective) relative to the original videos, and 50.4% (objective) and 47.7% (subjective) relative to the video remakes.

Lefebvre-Albaret (2011) conducted two sign language animation comprehension studies. For the first study, Deaf signers were presented with animations of isolated signs. Each animation was shown five times, after which the original video of a human signer performing the sign was presented. Comprehension was measured after each viewing. The average comprehension rate was 58% upon the first viewing of the animations and increased to 83% after three viewings, where it remained stable even after further viewings. The comprehension score for the original videos was 98%.

In their feedback for the first study, participants suggested slowing down the speed of signing and changing the point of view from front view to a slight rotation around the vertical axis. Hence, for the second study, Lefebvre-Albaret (2011) tested the comprehension of isolated signs with an adapted signing speed (reduction by 50%) and point of view (rotation of 20 degrees around vertical axis). The synthesized signs now received comprehension scores of 80% at first viewing among the Deaf participants. The scores rose up to 95% after five viewings. The corresponding scores for the videos of human signers averaged at 97%. The participants of the second study mentioned as issues hampering comprehension aspects related to movement and head orientation as well as missing mouthings and facial expressions.

The participants were also shown full sentences at a regular signing speed and point of view, both with and without non-manual information. For full sentences (at a regular speed, with front view), the comprehension rate was between 33% and 62% lower for the animations than for the videos of human signers performing the same utterances. At the sentence level, participants mentioned as the main comprehension barrier inaccurate facial expressions and mouthings, followed by a perceived lack of realism of the avatar along with imprecise movements and hand shapes. Quantitatively, the presence or absence of non-manual information did not affect comprehension greatly. However, the presence of mouthings did help participants to focus on the avatar's face. The researchers summarized that non-manual components contribute to comprehension "if they are accurate and synchronized with manual parameters" (Lefebvre-Albaret, 2011, p. 5).

Smith & Nolan (2015) tested both subjective and objective comprehension. They showed each animation twice and assessed comprehension after each viewing. Contrary to Huenerfauth et al. (2007), they found that the participants' subjective ratings of their comprehension was lower than the objective comprehension scores measured through comprehension questions (46% vs. 60%). The increase in score from the first to the second viewing was between 6% and 18%, depending on the content of the animations.

5.2.2 Acceptance

Kipp, Nguyen, Heloir, & Matthes (2011) carried out what is to date the most comprehensive sign language animation acceptance study. They conducted two focus group rounds and an online survey. As part of the focus group studies, a total of eight native signers of German Sign Language (*Deutsche Gebärdensprache*) (DGS) were presented with six avatars signing content in different languages: American Sign Language (ASL), British Sign Language (BSL), Finnish Sign Language, DGS, and International Sign (IS). The fact that the signers were asked to rate some avatars signing content in a language other than their first sign language poses a methodological issue (Ebling, 2013). The researchers explained it with the fact that content in DGS was not available for the different avatars they were interested in evaluating. Most of the avatars that were shown during the focus group study were fully synthesized (cf. Section 5.1). The participants were asked to discuss their strengths and weaknesses and vote on specific aspects.

The participants of the online survey (N=317) evaluated three of the six avatars presented in the focus groups, two fully synthesized ("Forest" and "Max" in Figure 5.5) and one hand-crafted ("DeafWorld" in Figure 5.5, corresponding to the Pedro avatar shown in Figure 5.2). The participants rated the avatars on a five-point scale with respect to different criteria, such as naturalness of movements, emotional expression, or degree of charisma. The results displayed in Figure 5.6 show that the hand-crafted avatar (dark bars) received more positive ratings than the two fully synthesized avatars (light bars).

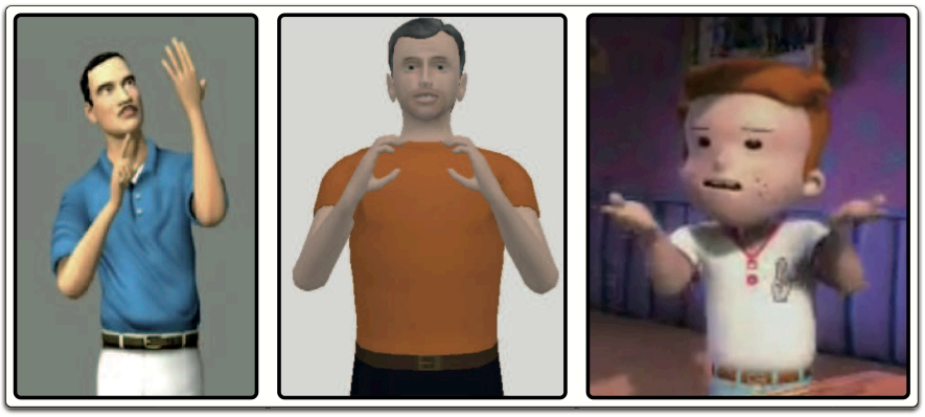


Figure 5.5: Avatars assessed in online survey: "Forest", "Max", and "DeafWorld" (figure from Kipp, Nguyen, et al., 2011)

This was reinforced in the focus group interviews, where the participants judged the fully synthesized avatars as being stiff and at times robot-like. In particular, they found

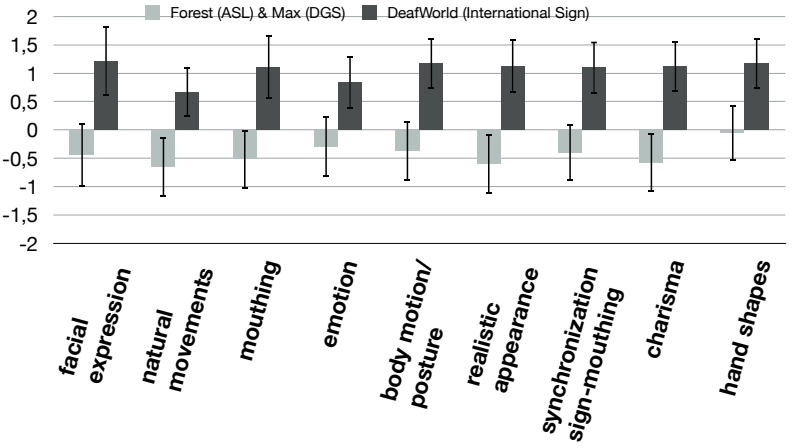


Figure 5.6: Results of sign language animation acceptance study (figure from Kipp, Nguyen, et al., 2011)

- that the avatars did not exhibit enough movement of head, shoulders, and torso;
- that they made too little use of the entire signing space;
- that their upper-body movements were not sufficiently smooth and relaxed;
- that they were missing variation in the movement of eyebrows, eyelids, and eyes;
- that they kept permanent eye contact in an obtrusive manner;
- that they used too few mouthings; and
- that, where mouthings were used, there was sometimes a mismatch between the duration of the manual activities and the mouthings.

Notably, the researchers observed an increase in acceptance through mere participation in the study: When asked the two questions “Do you think avatars are useful?” and “Do you think Deaf people would use avatars?” at the beginning and at the end of the study, the participants were significantly more in favor of avatars at the end of the study, both in the focus group interviews and in the online survey.

5.3 Synthesizing DSGS train announcements

The work reported here was in synthesizing the DSGS train announcements described in Section 3.6. This was the last step in the process of automatically translating German train announcements into synthesized DSGS as shown in Figure 4.1. As an animation system, I used JASigning.

5.3.1 JASigning

The Java Avatar Signing (JASigning) system (Glauert & Elliott, 2011; Jennings, Elliott, Kennaway, & Glauert, 2010; Kennaway, Glauert, & Zwitserlood, 2007) was developed during several international projects.[3] Its main release is freely available for research purposes[4] and offers different avatars, of which one is the Anna character shown in Figure 5.4. Other characters have been created for specific projects. The characters were built with frequently used 3D modelling software such as 3ds Max and Poser. Features relevant to sign language were added through the proprietary ARP Toolkit.[5]

In Section 3.6.3, Gestural Signing Gesture Markup Language (SiGML) was described, and reference was made to the fact that JASigning requires this variant of SiGML as input. Gestural SiGML code is received by the AnimGen animation engine (Kennaway et al., 2007) in JASigning, which generates motion data that can be used for the chosen avatar (in principle, for any avatar) in CAS format.[6] AnimGen applies an inverse kinematic approach. During this process, information that is necessary for the animation but not specified in the SiGML code has to be guessed. For example, if the SiGML code contains information about a contact between the avatar's hands, AnimGen has to guess the detailed nature of the contact (e.g., hands side by side vs. one above the other) based on heuristics (Kennaway et al., 2007).

Apart from SiGML code, AnimGen requires the input of four files defining the physical appearance of the avatar:

[3] An early version of the predecessor to JASigning, SiGMLSigning, was built with motion capturing techniques (cf. Section 5.1). The transition to notation-based synthesis occurred during a project named eSIGN.

[4] `http://vh.cmp.uea.ac.uk/index.php/JASigning` (last accessed December 17, 2015).

[5] `http://vh.cmp.uea.ac.uk/index.php/ARP` (last accessed December 17, 2015).

[6] Alternative output formats such as BVH or VRML are also supported.

1. a main avatar definition file containing
    - a list of vertices that make up the polygons of the avatar's surface mesh;
    - a link to a texture map defining the appearance of the avatar's skin and clothing;
    - a definition of the avatar's skeleton; and
    - information on how the surface mesh is attached to the skeleton;
2. an avatar standard description file;
3. an AnimGen configuration data file; and
4. a file controlling the non-manual features.

The file controlling the non-manual features contains mappings of SiGML attribute values (such as RB for raised eyebrows or NO for head nod, cf. Section 3.6.3) to morph targets, which are points on the facial mesh that may be deformed. Each morph target reference in the non-manual features file carries the attributes `name`, `amount`, and `timing`. The `amount` attribute specifies the amplitude of the morph, normally ranging between 0.0 and 1.0. The `timing` attribute consists of tags that control

- whether the morph is anchored to the start of the interval during which it is played;
- how long the attack time is;
- how the attack time is performed;
- how long the sustain time is;
- how long the release time is;
- how the release is performed; and
- whether the morph is anchored to the end of the interval during which it is played (Jennings et al., 2010).

The motion data generated by AnimGen specifies a sequence of frames, each of which is timestamped and contains information on the positions and (relative) rotations of the bones of the skeleton as well as on morph target amounts. Figure 5.7 shows motion data in CAS format for the sign LAUTSPRECHER ('LOUDSPEAKER') in DSGS.

```
<CAS version="2.1" avatar="anna">
  <frames count="84" signCount="1">
    <signStart index="0" gloss="LAUTSPRECHER"/>
    <frame index="0" isComplete="true" time="0"
    duration="20" boneCount="74" morphCount="51">
      <morph name="aaa" amount="0"/>
      <morph name="ooo" amount="0"/>
      <morph name="pout" amount="0"/>
      ...
      <bone name="ROOT">
        <position x="0" y="0" z="0"/>
        <qRotation x="0" y="0" z="0.7073" w="0.7069"/>
      </bone>
      <bone name="SPI1">
        <position x="0" y="0" z="0"/>
        <qRotation x="0" y="0" z="0" w="1"/>
      </bone>
      ...
    </frame>
  </frames>
</CAS>
```

Figure 5.7: Motion data for the sign LAUTSPRECHER ('LOUDSPEAKER') in DSGS

The motion data is rendered in real time by a conventional 3D character renderer (OpenGL). JASigning uses a Java binding for OpenGL (JOGL). Figure 5.8 visualizes the entire process of creating a sign language animation from Gestural SiGML code in JASigning.

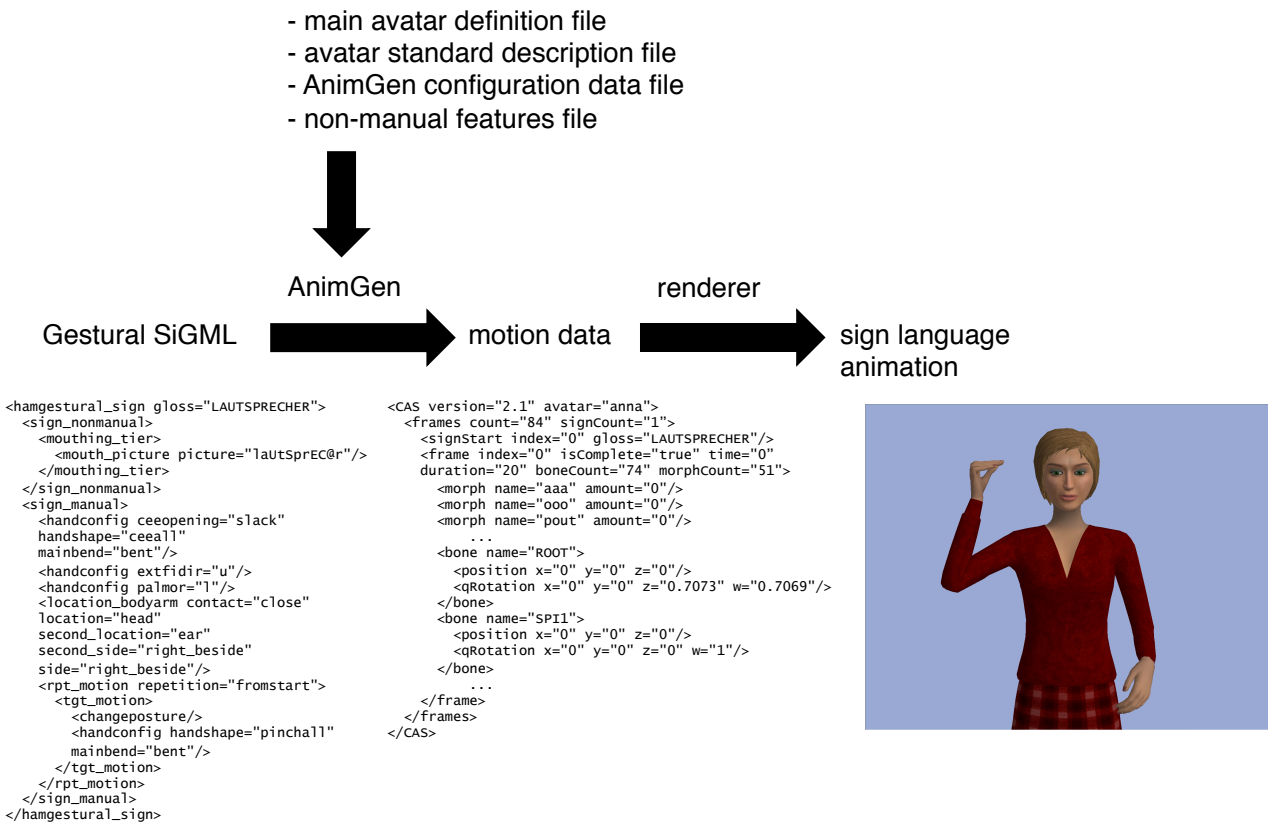


Figure 5.8: JASigning animation pipeline

5.3.2 Modifications to JASigning

Deliverables of the projects during which the JASigning system was developed and notes on a website are the main source of documentation for the system. Not all planned features have been fully implemented, some because they are used very infrequently, others because there is insufficient linguistic research on which to base an implementation. Together with my two Deaf collaborators, I identified the avatar functionality we needed for our project. In close collaboration with the developers of JASigning, I then found workarounds for those features that were not yet available in the system.

For example, the Gestural SiGML `<hamgestural_segment>` element (cf. Section 3.6.3), with which non-manual features can be applied to multiple signs, is not yet implemented. To replace its functionality, I modified the timing behavior of the non-manual features I wanted to extend over more than one sign.[7]

The `fitpicturetomanual`[8] attribute that synchronizes the duration of the mouthing and the manual activity of a sign is also not yet implemented. Substituting this attribute is not straightforward. In our case, the duration of a mouthing mostly exceeded the duration of the corresponding manual activity. Since the participants of our focus group study (cf. Section 5.3.3) remarked that the speed of the mouthings was generally too low, I sped up the mouthings by 20%.

In addition, I adjusted many of the SiGML-to-morph mappings: For example, I modified the code `SH` (head shake) in such a way that it involved fewer movements of the head with higher amplitudes. These changes were again motivated by feedback from Deaf experts.

While I was able to find workarounds for most features that were not yet available in the system, one remaining issue was how to cause non-manual components to slightly precede the manual activity of a sign. For example, the DSGS train announcements contained indexical (pointing) signs. The signs were accompanied by a shift in eye gaze towards the location of the indexical sign. In order for the signing to appear natural in this case, the onset of this non-manual component should precede the manual activity (pointing) slightly.[9]

[7] As outlined in Section 5.3.1, the morph(s) underlying a non-manual feature can be anchored to the start and/or end of a manual activity.

[8] Mouthings are sometimes referred to as "mouth pictures", in contrast to mouth gestures (cf. Section 2.1), which are also known as "mouth forms".

[9] This effect has also been observed for other sign languages (Braffort et al., 2013; McDonald et al., 2013).

5.3.3 Evaluation

Like Kipp, Nguyen, et al. (2011), I carried out a focus group study to obtain feedback from members of the DSGS community on how to improve the avatar signing DSGS train announcements. I chose the focus group method over single-case studies in order to provide a more informal setting, where Deaf people were free to exchange their thoughts rather than feel like they were part of a clinical experiment (Huenerfauth et al., 2007).

I followed the recommendation of Kipp, Nguyen, et al. (2011) to provide a sign-language-only setting, i.e., no hearing persons were allowed in the room in which the evaluation took place, myself included. One of the two Deaf members of our project acted as session moderator. We invited seven participants who were active members of the Deaf community and early learners of DSGS, which, different from Kipp, Nguyen, et al. (2011), I believe to be a crucial prerequisite for a successful evaluation. The group consisted of four men and three women of ages 22 to 69 (cf. Table 5.2 for the complete age distribution).

Participant ID	Age	Gender
1	22	F
2	39	M
3	42	M
4	49	F
5	51	F
6	58	M
7	69	M

Table 5.2: Demographic information about the participants of the study

The chairs were arranged in a semicircle, without table to help provide a more casual and personal atmosphere as well as assure that all participants could see both the screen and each other. One of the participants had Usher syndrome, i.e., he is Deaf as well as gradually becoming blind. Since he has difficulty adjusting to different lighting conditions and backgrounds, we placed one chair in front of a dark background. The moderator asked each participant wanting to make a statement to take a seat in this chair. Figure 5.9 shows the arrangement of seats. The discussion was recorded with four cameras (of which two are visible in Figure 5.9). Nine signed sentences were projected on a screen. The sentences had been chosen so as to reflect important characteristics of the sign language of our corpus, such as use of fingerspelling, time specifications, indexical signs, or lists of signs (cf. Section 3.6.1). For every sentence, the moderator asked for the participants' subjective opinion. She replayed avatar sequences upon request.

Figure 5.9: Focus group study setting

The participants recommended to slightly raise the avatar's eye gaze so that it would appear to be directed more towards the viewer. They found the posture of the avatar and the display window appropriate. However, they felt the transitions between some signs to be too abrupt. Moreover, they recommended for the hands to return to a neutral position at the end of every signed announcement rather than to come to rest in the final posture of the announcement.

The participants recommended slightly speeding up the mouthings. They also observed that the avatar's teeth and tongue were hardly visible; they found visibility to be necessary, e.g., when forming the mouthing for the fingerspelling sign -N-. They also found the speed of fingerspelling to be too high.

A long discussion evolved about how to deal with lists of place names. Where several place signs appeared together, we had introduced a short pause after each. The participants found that this was not sufficient. They discussed the following as different possible strategies:

- Preceding every place sign with the sign ORT ('PLACE') as a contextualization marker;
- Returning the hands to a neutral position after every place sign; or
- Performing a sign like THEMAWECHSEL ('CHANGE-OF-TOPIC') or WEGSCHIEBEN ('PUSH-ASIDE') after every place sign.

In the end, they opted for a combination of the first two strategies: performing the sign ORT once, then returning the hands to a neutral position after every place sign. The participants also suggested using the contextualization marker ORT together with single occurrences of place signs, even the widely known ones such as ZÜRICH, BASEL, or LUZERN.

Our conventions for time specifications had initially adhered to the format UHR <STUNDEN> PUNKT <MINUTEN> (‘CLOCK <HOUR-NUMBER> DOT <MINUTE-NUMBER>’) to reflect the fact that they originated in a timetable. However, the participants did not approve of this format. They suggested using instead a phrasing more familiar to them without the sign PUNKT (‘DOT’): <STUNDEN> UHR <MINUTEN> (‘<HOUR-NUMBER> CLOCK <MINUTE-NUMBER>’)

Regarding time specifications, the participants also remarked that a spatial offset between the signing location of the number of hours and the number of minutes was missing: They pointed out that in a temporal expression like 22 UHR 41 (‘22 CLOCK 41’), the number of hours (22) should be signed in front of the body and the succeeding number of minutes (41) slightly to the right. The same convention was recommended for train names involving numbers, e.g., S6, where S should be signed in front of the body and 6 slightly to the right.

The participants also found that the default transition time between specific combinations of signs was too long. This involved compound-like sign sequences such as BAHN VERKEHR (‘RAILROAD TRAFFIC’), ABFAHRT ORT (‘PLACE OF DEPARTURE’), or FAMILIE WAGEN (‘FAMILY WAGON’), but also cases in which DSGS uses two signs to refer to a single concept, like AUGE VORSICHT (‘EYE CAUTION’) for *Vorsicht* (‘caution’), VERSPÄTUNG NACH (‘DELAY AFTER’) for *Verspätung* (‘delay’), or SCHLIESSEN ZU (‘CLOSE CLOSED’) for *schliessen* (‘close’).

Following this feedback of the focus group participants, I made several improvements to the DSGS avatar. For example, I caused the hands to return to a neutral position at the end of every signed announcement. I slightly sped up the mouthings and decreased the speed of fingerspelling. I introduced the contextualization marker ORT (‘PLACE’) before place signs and for lists additionally caused the hands to return to a neutral position after every place sign. I also changed the format of time specifications such that different sets of glosses and corresponding HamNoSys notations for numbers were created: Instances of <STUNDEN> (‘<HOUR NUMBER>’) were signed in front of the signer’s body. For instances of <MINUTEN> (‘<MINUTE NUMBER>’), two cases were possible: If the number was between 00 and 09, the first digit was signed in front of the signer’s body and the second to the right; in all other cases (numbers from 10 onward), the digits were signed as one number to the right of the signer’s body. I also eliminated the temporal gap between compound-like sign sequences by introducing additional (compounded) lexicon entries for these occurrences.

5.4 Synthesizing the DSGS finger alphabet

While the work described in Section 5.3 was in producing animated DSGS train announcements, I also worked on synthesizing the DSGS finger alphabet (cf. Section 2.5). The animation system used for this was Paula, a hand-crafted avatar developed at DePaul University in Chicago. The remainder of this section is a short version of Ebling, Wolfe, et al. (2015) and represents joint work with the ASL Group at DePaul University.

Most existing tools for learning the finger alphabet of a sign language display one still image for each letter of a fingerspelling sequence. This is the case for the DSGS fingerspelling tutor interface shown in Figure 5.10: The fingerspelling sequence S-A-R-A-H is represented with five images.[10] In doing so, these tools do not account for all of the salient information inherent in fingerspelling: According to Wilcox (1992), when perceiving a fingerspelling sequence, the transitions between the letters are more important than the holds, i.e., more important than the canonical hand shapes of the letters. Transitions are usually not represented in sequences of still images.

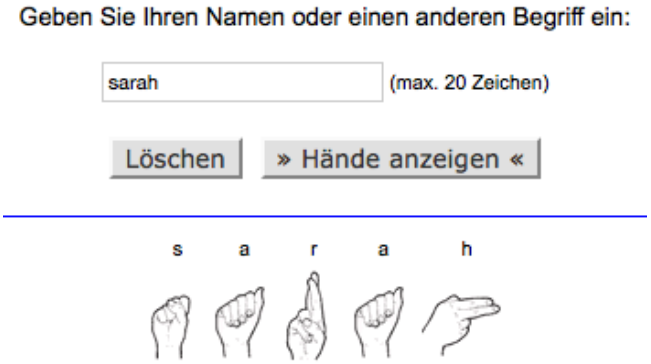


Figure 5.10: Fingerspelling tutor for DSGS

More recently, animation has been included in fingerspelling learning tools (Wolfe et al., 2006). This approach "has the flexibility to shuffle letters to create new words, as well as having the potential for producing the natural transitions between letters" (Toro, McDonald, & Wolfe, 2014, p. 561). The difference between an animation and a still-only representation of a fingerspelling sequence is shown schematically in Figure 5.11 for the example of the ASL sequence T-U-N-A: A still-only representation (upper row of Figure 5.11) would typically display four images for the

[10] `http://www.gebaerden-sprache.ch/informationen-zur-gebaerdensprache/fingeralphabet/fingeralphabet-mein-name/index.html` (last accessed October 1, 2015).

sequence under consideration, corresponding to the canonical form of each of the fingerspelling signs -T-, -U-, -N-, and -A-. By contrast, an animation (lower row of Figure 5.11) naturally also includes the transitions, as hinted at by the additional still images in the static capture of the animation.

Figure 5.11: Still images (above) vs. animation (below): Fingerspelling sequence T-U-N-A in ASL (figure from Wolfe et al., 2006)

To my knowledge, the application shown in Figure 5.10 is the only fingerspelling tutor for DSGS. Recall from Section 2.5 that use of the finger alphabet is more recent in DSGS than, e.g., in ASL. Together with the ASL Group at DePaul University, I synthesized the finger alphabet of DSGS as a first step towards a fingerspelling learning tool that employs sign language animation for this language.

5.4.1 Creating a set of synthesized DSGS hand postures and transitions

Synthesizing the DSGS manual alphabet consisted of producing hand postures (hand shapes with orientations) for each letter of the alphabet and transitions for each pair of letters such that postures and transitions could be seamlessly concatenated at runtime. The finger alphabet of DSGS was introduced in Figure 2.10 of Section 2.5. Recall that it features dedicated signs for -Ä-, -Ö-, and -Ü- as well as for -CH- and -SCH-.

Our work built on a previous system that synthesized the manual alphabet of ASL (Wolfe et al., 2006). Apart from the five additional signs just mentioned, the DSGS manual alphabet contains four hand shapes (those of -F-, -G-, -P-, and -T-) that are distinctly different from ASL. Further, the five letters -C-, -M-, -N-, -O-, and -Q- have similar hand shapes in ASL and DSGS but required smaller modifications, such as a change in orientation or adjustments in the fingers. Hence, overall, 14 out of the 30 hand postures of the DSGS finger alphabet needed modification from the ASL manual alphabet. All resulting hand postures were reviewed by native signers.

In creating the transitions, there was a potential of collisions between fingers when applying naive interpolation: For example, when transitioning from -N- to -A- in the ASL fingerspelling sequence T-U-N-A shown in Figure 5.11, the index and the middle finger have to lift first before the thumb can move outward to escape collision between the three fingers. For such cases, the ASL fingerspelling system contains transitional hand shapes that are inserted in-between two letters, forcing certain fingers to move before others to create clearance. Such a hand shape can be seen in the third frame from the right in the second row of Figure 5.11: As a first movement in the transition from -N- to -A-, the index and the middle finger lift.[11] The same was done for the DSGS fingerspelling system. Because of the overlap between the DSGS and ASL manual alphabets, along with the fact that most of the new or modified hand postures had hand shapes that were generally open (Brentari, 1998), it was possible to use the same set of transitional hand shapes for DSGS as for ASL.

5.4.2 Evaluation

We carried out a study to assess the comprehension of animated DSGS fingerspelling sequences produced from the set of hand postures and transitions described in Section 5.4.1. We conducted the study online using a remote testing system, *LimeSurvey*.[12] The remote approach has advantages over face-to-face testing in that it facilitates a large recruitment area and allows participants to complete the survey at any time. The survey was accessible from most web browsers and compatible across major operating systems. Any person with DSGS fingerspelling skills was invited to participate in the study. The call for participation was distributed via an online portal for the DSGS community[13] as well as through personal messages to persons known to fulfill the recruitment criteria.

Participants accessed the study through a URL provided to them. The first page of the website presented information about the study in DSGS (video of a human signer) and German (both text and video captions that represented a back-translation of the DSGS signing). Participants were informed of the purpose of the study, that participation was voluntary, that answers were anonymous, that items could be skipped, and that they could fully withdraw from the study at any

[11] Details of this method can be found in Wolfe et al. (2006).
[12] `https://www.limesurvey.org/en/` (last accessed October 1, 2015).
[13] `http://www.deafzone.ch/` (last accessed October 1, 2015).

time. Following this, they filled in a background questionnaire, which included questions about their hearing status, first language, preferred language, and age and manner of DSGS acquisition.

A detailed instruction page followed, on which the participants were informed that they were about to see 22 fingerspelled words signed by either a human or a signing avatar (the Paula avatar). Following this, they were told that their task was to type the letters of the word in a text box. Figure 5.12 shows a screenshot of the study interface for each of the two display modes. The videos of the human signer had been resized and cropped to match the animations.

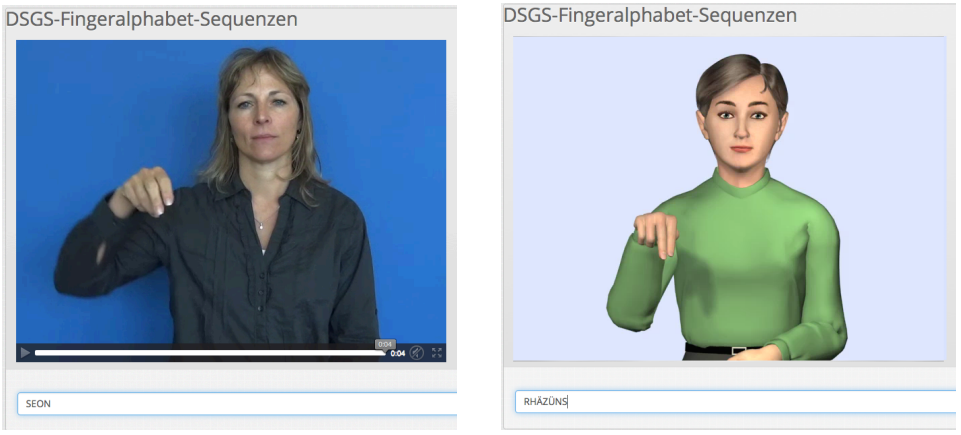


Figure 5.12: Online study interface (figures from Ebling, Wolfe, et al., 2015)

The participants were told that the fingerspelled words they were going to see were names of Swiss towns. An effort had been made to include only fingerspelled words (here: town names) for which no well-known lexical sign exists, in contrast to previous fingerspelling reception studies (Geer & Keane, 2014; Hanson, 1981). This was deemed an important prerequisite for a successful study. The items had been chosen based on the following criteria:

- They were names of towns with train stations that were among the least frequented based on a list obtained from the Swiss Federal Railways (*Schweizerische Bundesbahnen*) (SBB).
- The town names were of German or Swiss German origin.
- The town names in the resulting set of items varied with respect to their length (number of letters).
- In the resulting set of items, each letter of the DSGS finger alphabet occurred at least once (except for -X-, which did not occur in any of the town names that met all of the above criteria).

The 20 study items had an average length of seven letters, with a maximum of twelve (W-E-R-T-H-E-N-S-T-E-I-N) and a minimum of three (T-Ä-SCH). The study items were assigned to participants such that each item appeared as either a video of a human signer or an animation. Each participant saw ten videos and ten animations, and items were presented in random order. The study items were preceded by two practice items that were the same for all participants: The first was a video of a human signer fingerspelling S-E-O-N, the second an animation of R-H-Ä-Z-Ü-N-S (cf. Figure 5.12).

The human signer was a female native DSGS signer (Deaf-of-Deaf) who had been asked to sign at natural speed but without using mouthings. Recall from Section 2.5 that mouthings are common in DSGS, and that fingerspelling sequences in this language are typically accompanied by mouthings. We refrained from incorporating them into the animated fingerspelling sequences to obtain information about the comprehensibility of the actual fingerspelling. The fingerspelling rate of the human signer was 1.76 letters per second. The same rate was used for the animations. Note that it is below the minimum rate of 2.18 reported by Keane & Brentari (2015) (cf. Section 2.5), which, as in Section 5.3.3, points in the direction of a lower speed of fingerspelling in DSGS than in other sign languages.

The participants were informed that they could view a video as many times as they wanted. Limiting the number of viewings was felt to exert undue pressure. This also meant that there was no limit to the response time for an item. The response time was recorded as metadata.

Once participants had completed the main part of the study, they were asked to provide feedback on the following aspects:

- Appropriateness of the rate of fingerspelling;
- Comprehensibility of the individual letters and transitions between letters; and
- General feedback on the fingerspelling sequences shown.

On the final page, participants were thanked for their contribution and given the possibility to leave their e-mail address if they wanted to receive information on the results of the study. If provided, the e-mail address was not saved together with the rest of the data to ensure anonymity. All data was stored in a password-protected database.

The entire study was designed to take a maximum of 20 minutes to complete. This was assessed through a pilot study with three participants, in which the average time taken to complete the study was 17 minutes.

The study remained online for one week. During this time, 65 participants completed it, of which 31 were hearing, 24 Deaf, and six hard-of-hearing. Four participants indicated that they did not fall into the three categories proposed for hearing status, referring to themselves as "using sign and spoken language", "deafened", "CODA" (child of Deaf adult), and "residual hearing/profoundly hard-of-hearing". The average time taken to complete the entire survey was 20 minutes and 12 seconds.

For the 20 main study items (excluding the two practice items), 1 284 responses were submitted. In relation to the 1 300 possible responses (20 items $\times$ 65 participants), this meant that a total of 16 responses had been skipped.[14] They were treated as incorrect responses for the purpose of computing the comprehension rate.

For each of the 1 284 responses given, we automatically determined whether it was correct, ignoring umlaut expansions (*ä*→*ae*, etc.) and differences in case. Table 5.3 displays the comprehension rates: The mean percentage of correct responses was 93.91% for sequences fingerspelled by the human signer and 90.06% for sequences fingerspelled by the avatar. Also displayed are the binomial confidence intervals at a confidence level of 95%. They indicate a 95% confidence that the comprehension rate of the signing avatar is above 87.75% and below 92.37%. This result is highly satisfactory. It is visualized in Figure 5.13.

Display mode	**Comprehension rate (%)**	**Confidence interval lower bound (%)**	**Confidence interval upper bound (%)**
Human signer	93.91	92.05	95.76
Signing avatar	90.06	87.75	92.37

Table 5.3: Percentage of correct responses

Comprehension rates below 100% for human signing have been reported in previous studies, such as in Kipp, Heloir, & Nguyen (2011) and Lefebvre-Albaret (2011) (cf. Section 5.2.1). We hypothesize that in our case, the less-than-perfect comprehension scores for human signing were due at least partly to the fact that mouthings were absent from the fingerspellings. While this

[14] Recall that participants were given the option of not responding at any point in the study.

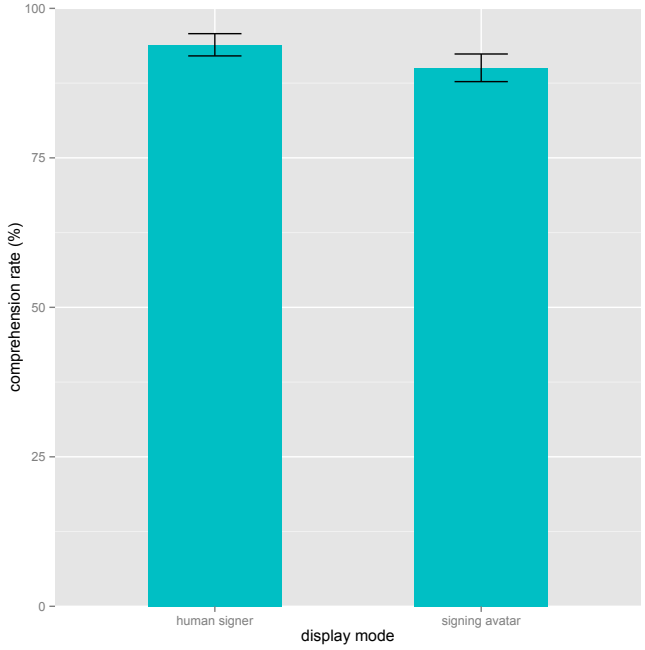


Figure 5.13: Percentage of correct responses: Human signer vs. avatar

was a methodological decision made to ensure that what was being measured was actual fingerspelling comprehension, several participants alluded to the lack of mouthings in the post-study questionnaire.

A comprehension rate of 100% was obtained for three sequences fingerspelled by the human signer (R-E-A-L-P, R-E-U-T-L-I-N-G-E-N, and S-E-D-R-U-N) and for three sequences produced by the signing avatar (B-E-V-E-R, H-U-R-D-E-N, and M-O-S-E-N).

To obtain information about individual letters that may have been hard to comprehend with the signing avatar, we performed a confusion analysis. The results showed that three letters were mistaken for other letters more often in sequences fingerspelled by the signing avatar than in sequences fingerspelled by the human signer: -F- (confused with -T- and -B-), -P- (confused with -G- and -H-), and -R- (confused with -U-). One letter, -H-, was confused more often in sequences fingerspelled by the human signer than in sequences fingerspelled by the signing avatar; it was mistaken with -G-, -L-, and -U-.

A confusion analysis between pairs of letters was also performed to obtain pointers to transitions that potentially needed improvement. Comprehension was lower for four transitions with the signing avatar than with the human signer: F-I (mistaken for T-I and B-I), L-P (mistaken for L-G and L-H), L-R (mistaken for L-U), and R-I (mistaken for U-I). This overlaps with the qualitative feedback in the post-study questionnaire that asked for letters and transitions that were

particularly hard to understand: Several participants mentioned the avatar's transitions into -G-, -I-, -P-, and -Q- as well as the transitions between -D- and -Q- and -L- and -P-. In addition, 12 out of 65 participants deemed the hand orientation of -Q- inaccurate.

In the general comments section, a number of participants remarked that the fingerspelling of the human signer was easier to understand than that of the signing avatar. Some participants noted that this was due to the hand appearing too small in the animations. At the same time, multiple participants commented on the quality of the signing avatar as being "surprisingly good". Repeated mention was made of the impression that short fingerspelling sequences were easier to understand than longer ones, regardless of whether they were signed by a human or an avatar.

One participant encouraged the introduction of speed controls for the signing avatar. In the post-study questionnaire rating of the speed of fingerspelling, the majority of the participants (number of responses: 62) deemed the speed appropriate (56.45%), followed by 35.48% who rated it as being too fast. 4.84% classified it as too slow, and 3.23% deemed it much too fast. No participant rated the speed as being much too slow. The numbers are summarized in Table 5.4.

Rating	**Responses (%)**
much too slow	0.00
too slow	4.84
appropriate	56.45
too fast	35.48
much too fast	3.23

Table 5.4: Speed of fingerspelling

5.5 Summary

This section has dealt with sign language animation. I have described the three conceptually different approaches to sign language animation, those being hand-crafted animation, animation based on motion capturing, and synthesis from form notation. Synthesis from form notation is the most flexible approach, since sign forms can easily be modified in context. Notation-based avatars have the potential to increase access to information for Deaf sign language users in cases where content is not persistent, e.g., on the web.

While machine translation research over the years has brought forth a number of metrics that allow for automatic evaluation of the output, no such metric exists for assessing the quality of

signing avatars. Sign language animation evaluation studies so far have been carried out in the form of user studies. Commonly assessed variables include comprehension and acceptance. However, the evaluation process is not standardized as of yet.

I have introduced the animation system I employed to create synthesized DSGS train announcements. I have described the modifications I made to the system and have reported on an evaluation of the DSGS avatar among the target community, as a result of which further changes were made to the avatar.

Finally, I have presented joint work with the ASL Group at DePaul University in synthesizing the finger alphabet of DSGS towards a fingerspelling learning tool for this language. The rationale behind using animation in such a tool is that animation is capable of representing the transitions between letters, which form a salient part of fingerspelling reception. I have reported on the process of creating a set of hand postures and transitions for DSGS as well as on the findings of a study assessing the comprehensibility of the resulting animations. The comprehension rate of the signing avatar was highly satisfactory. Information about individual letters and transitions that would benefit from improvement was also obtained.

Future work in the area of sign language animation might pursue two directions: firstly, that of improving the quality of signing avatars, and secondly, that of standardizing the process of evaluating the quality of signing avatars. Initial efforts have been made: For example, researchers have made available language-specific evaluation stimuli and associated comprehension and acceptance questions (Huenerfauth & Kacorri, 2014) and have proposed demographic and experiential questions to include in background questionnaires of evaluations (Kacorri, Huenerfauth, Ebling, Patel, & Willard, 2015). Work in this direction should continue.

Chapter 6

Conclusion and outlook

6.1 Conclusion

This thesis has presented my work in automatic translation from German to synthesized Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*) (DSGS). My research was set along one of the three automatic sign language processing pipelines introduced in the beginning and covered the three areas corpus linguistics, machine translation, and sign language animation.

I have discussed linguistic properties of sign languages, specifically considering those features that are relevant for an understanding of the prerequisites of automatically processing sign languages. I have shown that non-manual information is a salient part of signing. I have discussed different ways of providing a written record of signs, the primary distinction being between meaning-based and form-based notation systems. I have introduced the concept of iconicity and presented an ongoing study aimed at establishing the lexical similarity between German Sign Language (*Deutsche Gebärdensprache*) (DGS) and DSGS. The study takes into account not the general notion of iconic motivation but the more specific concept of image-producing techniques. I have also given an introduction to fingerspelling.

My research was connected to the use case of building a system that translates written German train announcements of the Swiss Federal Railways (*Schweizerische Bundesbahnen*) (SBB) into DSGS, the ultimate output consisting of a signing avatar. I have discussed the process of building a parallel corpus of German/DSGS train announcements together with two Deaf native DSGS

signers. This consisted of translating German train announcements into DSGS glosses and non-manual information, signing the DSGS announcements in front of a camera, and notating the form of the signs. For translation into DSGS glosses and non-manual information, we developed conventions to ensure consistency. An important part of building the parallel corpus consisted of including all information necessary for the sign language animation step in the DSGS side of the parallel corpus. This meant using a form notation from which motion data could subsequently be generated. The resulting parallel corpus of German/DSGS train announcements consisted of 2 986 announcement pairs, rendering it comparable in size to other parallel corpora built for use in sign language machine translation.

I have described my work in training a statistical machine translation system on the train announcement data. Since train announcements represent a limited domain that is standardized both with respect to grammar and vocabulary, the machine translation system produced highly satisfactory evaluation scores. Moreover, I have presented a solution for including non-manual information in an automatic sign language processing pipeline, something which had been omitted in most previous research. My solution scheduled the generation of non-manual information after the core machine translation task and viewed it as a sequence classification task. Hence, the generation of non-manual information was conceived as the task of labelling glosses (as representations of the manual activity of signs) with non-manual features. Sequence classification is a technique commonly used in the automatic processing of spoken languages. As far as I can see, my work was the first to apply it to sign languages. The experimental approaches outperformed the baselines (non-sequential classifiers) in all but two cases, emphasizing the benefit of sequence classification for the problem at hand. The results also underlined the potential of a cascaded approach, i.e., of using the output of one classifier as additional input for another. In particular, they suggested that for DSGS, head information is more valuable for predicting eyebrow information than vice versa.

I have introduced three approaches to sign language animation and have shown that synthesis from form notation is the most flexible approach. I have reported on my work in synthesizing DSGS train announcements. I used the JASigning system, which supports synthesis from form notation. I modified the system to fit the needs of DSGS animation and performed an evaluation of the resulting DSGS avatar among the target community. Based on this evaluation, I made further changes to the DSGS avatar.

Departing from my work in translating written German train announcements into synthesized

DSGS as presented in this thesis, a similar system is currently being developed for the translation of French train announcements into Swiss French Sign Language (*Langue des Signes Française Suisse*) (LSF-CH) (Rayner et al., 2015).

I have also presented joint work in synthesizing the finger alphabet of DSGS with a view to developing a fingerspelling learning tool for this language. Using animations instead of still images in such a tool is an obvious choice, since animations are capable of rendering the transitions between letters that form a salient part of fingerspelling reception. I have reported on the process of creating a set of hand postures and transitions for DSGS as well as on the results of a study assessing the comprehensibility of the animations.

The result of my research is the first parallel corpus, machine translation system, and avatar for DSGS. Consequently, this thesis has established the preconditions of successful automatic processing of DSGS. Moreover, it has outlined the prerequisites for bridging the gap between sign language machine translation and sign language animation. It has also presented a solution for automatically generating non-manual information. Since this solution is based on machine learning, it can be applied to other sign languages as well. Lastly, the thesis has made a contribution to quality improvement of signing avatars.

6.2 Outlook

As has been shown throughout this thesis, acquiring sign language data is a heavily time-consuming task, which explains why many sign languages are low-resource languages. Corpora of the size of the one used for this thesis are large enough if they stem from a restricted domain such as that of train announcements. More generally speaking, current data-driven automatic sign language processing systems are successful if they operate on data that is inherently parametrized in one way or another. If these systems are to work well on more variable domains, more data is needed. Here, the potential of sign language recognition to speed up the (an)notation process could be leveraged. With this, form notations could be produced based on occurrences of signs in context instead of citation forms of signs, which would allow for capturing possible coarticulation effects. Moreover, partly automating the notation process would make it possible to record a wider variety of non-manual features than were considered for the work reported in this thesis.

Progress in sign language machine translation might include extending the functionality of these systems such that they are capable of dealing with a wider variety of phenomena found in sign

languages. Future work could also be directed towards developing a machine translation evaluation metric that takes into account the multi-level nature of sign languages. This becomes important if the representations of sign language in such translation systems are not merely strings of glosses.

Future work in automatic generation of non-manual information through sequence classification could look into dealing with non-manual information that is not predictable on the basis of glosses alone, i.e., cases for which information from the original source side is needed in addition.

With regard to sign language animation, the quality of signing avatars might steadily be improved. As was shown in this thesis, current signing avatars are often still perceived as stiff and unnatural. This is especially true for fully synthesized avatars. Because of their flexibility, these avatars have the largest application potential when it comes to providing access to information for Deaf persons in everyday life. Increasing the quality of these avatars also involves carrying out further research on the linguistic structure of sign languages. Future research in this area might also be concerned with standardizing the evaluation process.

References

Allwood, J. (2009). Multimodal corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 207–225). Berlin, Germany: De Gruyter Mouton.

Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.

Boyes Braem, P. (1983). Studying Swiss German Sign Language dialects. In *Proceedings of the 3rd International Symposium on Sign Language Research (SLR)* (pp. 247–253). Rome, Italy.

Boyes Braem, P. (1995). *Einführung in die Gebärdensprache und ihre Erforschung* (3rd ed.). Hamburg, Germany: Signum.

Boyes Braem, P. (2001a). Functions of the mouthing component in the signing of Deaf early and late learners of Swiss German Sign Language. In D. Brentari (Ed.), *Foreign vocabulary in sign languages: A cross-linguistic investigation of word formation* (pp. 1–47). Mahwah, NJ: Erlbaum.

Boyes Braem, P. (2001b). Functions of the mouthings in the signing of Deaf early and late learners of Swiss German Sign Language (DSGS). In P. Boyes Braem & R. Sutton-Spence (Eds.), *The hands are the head of the mouth: The mouth as articulator in sign languages* (pp. 99–133). Hamburg, Germany: Signum.

Boyes Braem, P. (2001c). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, *4*(1/2), 133–143.

Boyes Braem, P. (2005). *Gebärdensprachkurs Deutschschweiz, Stufe 4: Linguistischer Kommentar*. Zurich, Switzerland: GS-Media. (CD-ROM)

Boyes Braem, P. (2012a). Evolving methods for written representations of signed languages of the Deaf. In A. Ender, A. Leemann, & B. Waelchli (Eds.), *Methods in contemporary linguistics* (pp. 411–438). Berlin, Germany: De Gruyter Mouton.

Boyes Braem, P. (2012b). *Overview of research on signed languages of the Deaf.* (Lecture held at the University of Basel. Retrieved from `http://www.signlangcourse.org` (last accessed November 13, 2015))

Boyes Braem, P. (2014). Lautlos über alles sprechen und alles verstehen: Gebärdensprache. In E. Glaser, A. Kolmer, M. Meyer, & E. Stark (Eds.), *Sprache(n) verstehen* (pp. 59–84). Zurich, Switzerland: vdf-Hochschulverlag.

Boyes Braem, P., Groeber, S., Stocker, H., & Tissi, K. (2012). Weblexikon für Fachbegriffe in Deutschschweizerischer Gebärdensprache (DSGS) und Deutsch. *eDITion: Fachzeitschrift für Terminologie*, *2*, 8–14.

Boyes Braem, P., Haug, T., & Shores, P. (2012). Gebärdenspracharbeit in der Schweiz: Rückblick und Ausblick. *Das Zeichen*, *90*, 58–74.

Braffort, A., Filhol, M., Delorme, M., Bolot, L., Choisier, A., & Verrecchia, C. (2013). KAZOO: A sign language generation platform based on production rules. In *Proceedings of the 3rd International Symposium on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., … Roossin, P. (1990). A statistical approach to machine translation. *Computational Linguistics*, *16*(2), 79–85.

Bungeroth, J., Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A., & van Zijl, L. (2008). The ATIS sign language corpus. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)* (pp. 2943–2946). Marrakech, Morocco.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)* (pp. 249–256). Trento, Italy.

Chen, S. F., & Goodman, J. (1996). An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 310–318). Santa Cruz, CA.

Chiang, D. (2005). A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 263–270). Ann Arbor, MI.

Coerts, J. (1994). Constituent order in Sign Language of the Netherlands and the functions of orientations. In I. Ahlgren, B. Bergman, & M. Brennan (Eds.), *Perspectives on sign language structure* (pp. 69–88). Durham, UK: International Sign Linguistics Association.

Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002). Tessa, a system to aid communication with deaf people. In *Proceedings of the 5th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 205–212). Edinburgh, Scotland.

Crasborn, O. (2006). Nonmanual structures in sign language. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., Vol. 8, pp. 668–672). Oxford, UK: Elsevier.

Crasborn, O., & Hanke, T. (2003). *Additions to the IMDI metadata set for sign language corpora* (Tech. Rep.). ECHO project. (Retrieved from `http://sign-lang.ruhosting.nl/echo/docs/SignMetadata_Oct2003.pdf` (last accessed November 13, 2015))

Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., & Bergman, B. (2007). Sharing sign language data online: Experiences from the ECHO project. *International Journal of Corpus Linguistics*, *12*(4), 535–562.

Crasborn, O., & Zwitserlood, I. (2008). The Corpus NGT: An online corpus for professionals and laymen. In *Proceedings of the 3rd LREC Workshop on Representation and Processing of Sign Languages* (pp. 44–49). Marrakech, Morocco.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the 2nd International Conference on Human Language Technology Research (HLT)* (pp. 138–145). San Diego, CA.

Dreuw, P., & Ney, H. (2008). Towards automatic sign language annotation for the ELAN tool. In *Proceedings of the 3rd LREC Workshop on Representation and Processing of Sign Languages* (pp. 50–53). Marrakech, Morocco.

Ebling, S. (2010). *Generalized templates in example-based machine translation* (Unpublished master's thesis). University of Zurich.

Ebling, S. (2013). Evaluating a Swiss German Sign Language avatar among the Deaf community. In *Proceedings of the 3rd International Symposium on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL. (Retrieved from `http://www.zora.uzh.ch/85717/1/CAMERA_READY_sltat2013_submission_14.pdf` (last accessed November 20, 2015))

Ebling, S. (2016). Building a parallel corpus of German/Swiss German Sign Language train announcements. *International Journal of Corpus Linguistics (IJCL)*, *21*(1), 115–129.

Ebling, S., & Glauert, J. (2013). Exploiting the full potential of JASigning to build an avatar signing train announcements. In *Proceedings of the 3rd International Symposium on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL. (Retrieved from `http://www.zora.uzh.ch/85716/1/CAMERA_READY_sltat2013_submission_13.pdf` (last accessed November 20, 2015))

Ebling, S., & Glauert, J. (2015). Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society*, 1–11. (Retrieved from `http://dx.doi.org/10.1007/s10209-015-0408-1` (last accessed November 20, 2015))

Ebling, S., & Huenerfauth, M. (2015). Bridging the gap between sign language machine translation and sign language animation using sequence classification. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. Dresden, Germany. (Retrieved from `http://www.slpat.org/slpat2015/papers/ebling-huenerfauth.pdf` (last accessed November 20, 2015))

Ebling, S., Konrad, R., Boyes Braem, P., & Langer, G. (2015). Factors to consider when making lexical comparisons of sign languages: Notes from an ongoing study comparing German Sign Language and Swiss German Sign Language. *Sign Language Studies*, *16*(1), 30–56.

Ebling, S., Wolfe, R., Schnepp, J., Baowidan, S., McDonald, J., Moncrief, R., … Tissi, K. (2015). Synthesizing the finger alphabet of Swiss German Sign Language and evaluating the comprehensibility of the resulting animations. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. Dresden, Germany. (Retrieved from `http://www.slpat.org/slpat2015/papers/ebling-wolfe-et-al.pdf` (last accessed November 20, 2015))

Efthimiou, E., Fotinea, S., Vogler, C., Hanke, T., Glauert, J., Bowden, R., … Segouat, J. (2009). Sign language recognition, generation, and modelling: A research effort with applications in

Deaf communication. In C. Stephanidis (Ed.), *Universal access in human-computer interaction* (pp. 21–30). Berlin, Germany: Springer.

Efthimiou, E., Fotinea, S.-E., Hanke, T., Glauert, J., Bowden, R., Braffort, A., … Lefebvre-Albaret, F. (2012). The Dicta-Sign Wiki: Enabling web communication for the Deaf. In *Proceedings of the 13th International Conference on Computers Helping People with Special Needs (ICCHP)* (pp. 205–212). Linz, Austria.

Elliott, R., Glauert, J., Kennaway, R., & Marshall, I. (2000). The development of language processing support for the ViSiCAST project. In *Proceedings of the 4th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 101–108). Arlington, VA.

Erlenkamp, S. (2001). Lexikalische Klassen und syntaktische Kategorien in der deutschen Gebärdensprache: Warum das Vorhandensein von Verben nicht unbedingt Nomen erfordert. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000: Theorie und Anwendung* (pp. 67–91). Hamburg, Germany: Signum.

Estrella, P. (2008). *Evaluating machine translation in context: Metrics and tools* (Unpublished doctoral dissertation). University of Geneva.

Federico, M., & Cettolo, M. (2007). Efficient handling of n-gram language models for statistical machine translation. In *Proceedings of the 2nd ACL Workshop on Statistical Machine Translation* (pp. 88–95). Prague, Czech Republic.

Fischer, S. D. (1975). Influences on word order change in ASL. In C. Li (Ed.), *Word order and word order change* (pp. 1–25). Austin, TX: University of Texas Press.

Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J., & Ney, H. (2012). RWTH-PHOENIX-Weather: A large vocabulary sign language recognition and translation corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)* (pp. 3785–3789). Istanbul, Turkey.

Forster, J., Schmidt, C., Koller, O., Bellgardt, M., & Ney, H. (2014). Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)* (pp. 1911–1916). Reykjavík, Iceland.

Frishberg, N. (1979). Historical change: From iconic to arbitrary. In E. Klima & U. Bellugi (Eds.), *The signs of language* (pp. 67–84). Cambridge, MA: Harvard University Press.

Geer, L., & Keane, J. (2014). Exploring factors that contribute to successful fingerspelling comprehension. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)* (pp. 1905–1910). Reykjavík, Iceland.

Glauert, J. (2013). *Animating sign language for Deaf people.* (Lecture held at the University of Zurich, October 9, 2013 (unpublished))

Glauert, J., & Elliott, R. (2011). Extending the SiGML notation: A progress report. In *Proceedings of the 2nd International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*. Dundee, Scotland. (Retrieved from `http://vhg.cmp.uea.ac.uk/demo/SLTAT2011Dundee/8.pdf` (last accessed November 20, 2015))

Gutjahr, A. (2006). *Lesekompetenz Gehörloser: Ein Forschungsüberblick* (Unpublished master's thesis). University of Hamburg.

Hanke, T. (2001). *ViSiCAST Deliverable D5-1: Interface definitions* (Tech. Rep.). ViSiCAST project. (Retrieved from `http://www.visicast.cmp.uea.ac.uk/Papers/ViSiCAST_D5-1v017rev2.pdf` (last accessed November 20, 2015))

Hanke, T. (2013). *Corpus linguistics on sign language.* (Presentation given at the conference "From Hand to Mouth: A dialogue between signed and spoken language research", University of Zurich, September 6, 2013 (unpublished))

Hanke, T., Matthes, S., Regen, A., & Worseck, S. (2012). Where does a sign start and end? Segmentation of continuous signing. In *Proceedings of the 5th LREC Workshop on the Representation and Processing of Sign Languages* (pp. 69–74). Istanbul, Turkey.

Hanke, T., & Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)* (pp. 64–67). Marrakech, Morocco.

Hanson, V. (1981). When a word is not the sum of its letters: Fingerspelling and spelling. In *Proceedings of the 3rd National Symposium on Sign Language Research and Teaching* (pp. 176–185). Boston, MA.

Hemphill, C., Godfrey, J., & Doddington, G. (1990). The ATIS spoken language systems pilot corpus. In *Proceedings of the DARPA Speech and Natural Language Workshop* (pp. 96–101). Somerset, PA.

Himmelmann, N. (1998). Documentary and descriptive linguistics. *Linguistics*, *36*, 161–195.

Hoang, H. (2007). Factored translation models. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 868–876). Prague, Czech Republic.

Hong, S.-E., Hanke, T., König, S., Konrad, R., Langer, G., & Rathmann, C. (2009). *Elicitation materials and their use in sign language linguistics.* (Poster presented at the Sign Language Corpora: Linguistic Issues Workshop, London, UK. Retrieved from `http://www.bslcorpusproject.org/wp-content/uploads/posterlondonelizitation.pdf` (last accessed November 20, 2015))

Huenerfauth, M. (2003). *A survey and critique of American Sign Language natural language generation and machine translation systems* (Tech. Rep.). University of Pennsylvania. (Retrieved from `http://eniac.cs.qc.edu/matt/pubs/huenerfauth-2003-ms-cis-03-32-asl-nlg-mt-survey.pdf` (last accessed November 20, 2015))

Huenerfauth, M., & Hanson, V. (2009). Sign language in the interface: Access for Deaf signers. In C. Stephanidis (Ed.), *Handbook of universal access.* Boca Raton, FL: CRC Press.

Huenerfauth, M., & Kacorri, H. (2014). Release of experimental stimuli and questions for evaluating facial expressions in animations of American Sign Language. In *Proceedings of the 6th LREC Workshop on the Representation and Processing of Sign Languages* (pp. 71–76). Reykjavík, Iceland.

Huenerfauth, M., Zhao, L., Gu, E., & Allbeck, J. (2007). Evaluating American Sign Language generation through the participation of native ASL signers. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 211–218). Tempe, AZ.

Jennings, V., Elliott, R., Kennaway, R., & Glauert, J. (2010). Requirements for a signing avatar. In *Proceedings of the 4th LREC Workshop on the Representation and Processing of Sign Languages* (pp. 133–136). La Valetta, Malta.

Johnston, T. (1989). *Auslan: The sign language of the Australian Deaf community* (Unpublished doctoral dissertation). University of Sydney.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In *Proceedings of the 3rd LREC Workshop on Representation and Processing of Sign Languages* (pp. 82–87). Marrakech, Morocco.

Johnston, T., & Schembri, A. (1999). On defining lexeme in a signed language. *Sign Language & Linguistics*, *2*(2), 115–185.

Johnston, T., & Schembri, A. (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge, UK: Cambridge University Press.

Kacorri, H., Huenerfauth, M., Ebling, S., Patel, K., & Willard, M. (2015). Demographic and experiential factors influencing acceptance of sign language animation by Deaf users. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 147–154). Lisbon, Portugal.

Kacorri, H., Lu, P., & Huenerfauth, M. (2013). Effect of displaying human videos during an evaluation study of American Sign Language animation. *ACM Transactions on Accessible Computing*, *5*(2), 4:1–4:31.

Keane, J., & Brentari, D. (2015). Fingerspelling: Beyond handshape sequences. In M. Marschark & P. Spencer (Eds.), *The Oxford Handbook of Deaf Studies in Language: Research, policy, and practice* (pp. 146–160). New York, NY: Oxford University Press.

Kennaway, R., Glauert, J., & Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, *14*(3), 15:1–15:29.

Kipp, M., Heloir, A., & Nguyen, Q. (2011). Sign language avatars: Animation and comprehensibility. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents (IVA)* (pp. 113–126). Reykjavík, Iceland.

Kipp, M., Nguyen, Q., Heloir, A., & Matthes, S. (2011). Assessing the Deaf user perspective on sign language avatars. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 107–114). Dundee, Scotland.

Klima, E., & Bellugi, U. (1979). *Signs of language*. Cambridge, MA: Harvard University Press.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the 10th Machine Translation Summit* (pp. 79–86). Phuket, Thailand.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., … Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 177–180). Prague, Czech Republic.

Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)* (pp. 48–54). Edmonton, Canada.

Konrad, R. (2010). *Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010).* (Retrieved from `http://www.sign-lang.uni-hamburg.de/projekte/mfl/konrad_2010_fachgeblexika.pdf` (last accessed November 13, 2015))

Konrad, R. (2011). *Die lexikalische Struktur der Deutschen Gebärdensprache im Spiegel empirischer Fachgebärdenlexikographie: Zur Integration der Ikonizität in ein korpusbasiertes Lexikonmodell*. Tübingen, Germany: Gunter Narr Verlag.

Konrad, R. (2013). The lexical structure of German Sign Language (DGS) in the light of empirical LSP lexicography: On how to integrate iconicity in a corpus-based lexicon model. *Sign Language & Linguistics*, *16*(1), 111–118.

Labov, W. (1972). *Sociolingustic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Langer, G. (2005). Bilderzeugungstechniken in der Deutschen Gebärdensprache. *Das Zeichen*, *70*, 254–270.

Lavergne, T., Cappé, O., & Yvon, F. (2010). Practical very large scale CRFs. In *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 504–513). Uppsala, Sweden.

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8–29). London, UK: Longman.

Leeson, L., & Saeed, J. (2012). Word order. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook.* Berlin, Germany: De Gruyter Mouton.

Leeson, L., Saeed, J., Macduff, A., Byrne-Dunne, D., & Leonard, C. (2006). Moving heads and moving hands: Developing a digital corpus of Irish Sign Language. In *Proceedings of the Information Technology and Telecommunications Conference* (pp. 33–43). Carlow, Ireland.

Lefebvre-Albaret, F. (2011). *DictaSign Deliverable D7.4: Prototype evaluation synthesis* (Tech. Rep.). DictaSign project. (Retrieved from `http://cordis.europa.eu/docs/projects/cnect/5/231135/080/deliverables/001-DICTASIGNDeliverableD74FINALAres2012696081.pdf` (last accessed November 20, 2015))

Lefebvre-Albaret, F., Gibet, S., Turki, A., Hamon, L., & Brun, R. (2013). Overview of the Sign3D project: High-fidelity 3D recording, indexing and editing of French Sign Language content. In *Proceedings of the 3rd International Symposium on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL. (Retrieved from `https://hal.archives-ouvertes.fr/hal-00914661/en` (last accessed December 16, 2015).)

Lemnitzer, L., & Zinsmeister, H. (2006). *Korpuslinguistik: Eine Einführung*. Tübingen: Narr.

Lewis, M. P. (2009). *Ethnologue: Languages of the world* (16th ed.). Dallas, TX: SIL International. (Retrieved from `http://www.ethnologue.com/` (last accessed November 13, 2015))

Massó, G., & Badia, T. (2010). Dealing with sign language morphemes in statistical machine translation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)* (pp. 154–157). La Valetta, Malta.

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., & Berke, L. (2013). Toward lifelike animations of American Sign Language: Achieving natural motion from the Movement-Hold Model. In *Proceedings of the 3rd International Symposium on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh, Scotland: Edinburgh University Press.

McKee, R. (2015). *New Zealand Sign Language: A reference grammar*. Wellington, New Zealand: Bridget Williams Books. (E-book)

Mercer, K. C. R. (1993). Introduction to the Special Issue on Computational Linguistics using Large Corpora. *Computational Linguistics*, *19*(1), 1–24.

Mitchell, R., & Karchmer, M. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, *4*(2), 138–163.

Morgan, G., & Woll, B. (2002). The development of complex sentences in British Sign Language. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition: Trends in language acquisition research* (pp. 255–276). Amsterdam, Netherlands: John Benjamins.

Morrissey, S. (2008). *Data-driven machine translation for sign languages* (Unpublished doctoral dissertation). Dublin City University.

Neidle, C. (2002). *SignStreamTM annotation: Conventions used for the American Sign Language Linguistic Research Project: Report no. 11 American Sign Language Linguistic Research Project. ASLLRP annotation schema version 2.5* (Tech. Rep.). Boston University. (Retrieved from `http://www.bu.edu/asllrp/asllrpr11.pdf` (last accessed November 20, 2015))

Neidle, C. (2007). *SignStreamTM annotation: Addendum to conventions used for the American Sign Language Linguistic Research Project. Report no. 13 American Sign Language Linguistic Research Project. ASLLRP annotation schema version 3.0* (Tech. Rep.). Boston University. (Retrieved from `http://www.bu.edu/asllrp/asllrpr13.pdf` (last accessed November 20, 2015))

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., & Lee, R. G. (2001). *The syntax of American Sign Language: Functional categories and hierarchical structure* (2nd ed.). Cambridge, MA: MIT Press.

Nießen, S., Och, F., Leusch, G., & Ney, H. (2000). An evaluation tool for machine translation: Fast evaluation for MT research. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC)* (pp. 39–45). Athens, Greece.

Nishio, R., Hong, S. E., König, S., Konrad, R., Langer, G., Hanke, T., & Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) corpus project. In *Proceedings of the 4th LREC Workshop on the Representation and Processing of Sign Languages* (pp. 178–185). La Valetta, Malta.

Padden, C. (1988). *Interaction of morphology and syntax in American Sign Language*. New York, NY: Garland Press.

Padden, C., & Gunsauls, D. C. (2003). How the alphabet came to be used in a sign language. *Sign Language Studies*, *4*(1), 10–33.

Palfreyman, N., Sagara, K., & Zeshan, U. (2015). Methods in carrying out language typological research. In E. Orfanidou, B. Woll, & G. Morgan (Eds.), *Research methods in sign language studies: A practical guide* (pp. 173–192). Chichester, UK: Wiley-Blackwell.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 311–318). Philadelphia, PA.

Pfau, R., & Quer, J. (2010). Nonmanuals: Their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign languages* (pp. 381–403). New York, NY: Cambridge University Press.

Pizzuto, E., Rossini, P., & Russo, T. (2006). Representing signed languages in written form: Questions that need to be posed. In *Proceedings of the 2nd LREC Workshop on Representation and Processing of Sign Languages* (pp. 1–6). Genoa, Italy.

Poizner, H., Klima, E., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: MIT Press.

Portele, T., Krämer, J., & Stock, D. (1995). Symbolverarbeitung im Sprachsynthesesystem Hadifix. In *Proceedings of the 6th Conference Elektronische Sprachsignalverarbeitung* (pp. 97–104). Wolfenbüttel, Germany.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. (1989). *HamNoSys: Version 2.0: An introductory guide*. Hamburg, Germany: Signum.

Rayner, M., Armando, A., Bouillon, P., Ebling, S., Gerlach, J., Halimi, S., … Tsourakis, N. (2015). Helping domain experts build phrasal speech translation systems. In *Proceedings of the Future and Emerging Trends in Language Technology Workshop (FETLT)*. Seville, Spain. (Retrieved from `http://dx.doi.org/10.5167/uzh-115013` (last accessed December 16, 2015).)

Reilly, J., & Anderson, D. (2002). FACES: The acquisition of non-manual morphology in ASL. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 159–181). Amsterdam, Netherlands: John Benjamins.

Roth, L., & Clematide, S. (2014). Tagging complex non-verbal German chunks with Conditional Random Fields. In *Proceedings of the 12th Konvens Conference* (pp. 48–57). Hildesheim, Germany.

Sáfár, E., & Glauert, J. (2012). Computer modelling. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 1075–1101). Berlin, Germany: De Gruyter Mouton.

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.

Sang, E. F. T. K., & Veenstra, J. (1999). Representing text chunks. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL)* (pp. 173–179). Bergen, Norway.

San-Segundo, R., Lopez, V., Martin, R., Sanchez, D., & Garcia, A. (2010). Language resources for Spanish–Spanish Sign Language (LSE) translation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)* (pp. 208–211). La Valetta, Malta.

Segouat, J. (2010). *Modélisation de la coarticulation en Langue des Signes Française pour la diffusion* (Unpublished doctoral dissertation). Université Paris Sud.

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford, UK: Oxbow Books. (Retrieved from `http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm` (last accessed November 26, 2015))

Smith, R., & Nolan, B. (2015). Emotional facial expressions in synthesised sign language avatars: A manual evaluation. *Universal Access in the Information Society*, 1–10. (Retrieved from `http://dx.doi.org/10.1007/s10209-015-0410-7` (last accessed November 20, 2015))

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of Translation Edit Rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA)* (pp. 223–231). Cambridge, MA.

Stein, D., Forster, J., Zelle, U., Dreuw, P., & Ney, H. (2010). RWTH-Phoenix: Analysis of the German Sign Language corpus. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)* (pp. 225–230). La Valetta, Malta.

Stein, D., Schmidt, C., & Ney, H. (2012). Analysis, preparation, and optimization of statistical sign language machine translation. *Machine Translation*, *26*(4), 325–357.

Stokoe, W. (1960). Sign language structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics: Occasional Papers*, *8*, 1–41.

Stroppa, N., & Way, A. (2006). MaTrEx: DCU machine translation system for IWSLT 2006. In *Proceedings of the 3rd International Workshop on Spoken Language Translation (IWSLT)* (pp. 31–36). Kyoto, Japan.

Su, H.-Y., & Wu, C.-H. (2009). Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory. *IEEE Transactions on Audio, Speech and Language Processing*, *17*(7), 1305–1315.

Su, S., & Tai, J. (2009). Lexical comparison of signs from Taiwan, Chinese, Japanese, and American Sign Languages: Taking iconicity into account. In J. H.-Y. Tai & J. Tsay (Eds.), *Taiwan Sign Language and beyond* (pp. 149–176). Taiwan: The Taiwan Institute for the Humanities.

Sutton, C., & McCallum, A. (2012). An introduction to Conditional Random Fields. *Foundations and Trends in Machine Learning*, *4*(4), 267–373.

Sutton, V. (2010). *The International SignWriting Alphabet: SignWriting Alphabet Manual, ISWA 2010.* (Retrieved from `http://www.signwriting.org/archive/docs7/sw0636_SignWriting_Alphabet_Manual_2010.pdf` (last accessed November 13, 2015))

Sze, F. (2003). Word order of Hong Kong Sign Language. In A. Baker, B. van den Bogaerde, & O. Crasborn (Eds.), *Cross-linguistic perspectives in sign language research: Selected papers from TISLR 2000* (pp. 163–191). Hamburg, Germany: Signum.

Tillmann, C., Vogel, S., Ney, H., Zubiaga, A., & Sawaf, H. (1997). Accelerated DP-based search for statistical translation. In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech)* (pp. 2667–2670). Rhodes, Greece.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Netherlands: John Benjamins.

Toro, J. A., McDonald, J., & Wolfe, R. (2014). Fostering better Deaf/hearing communication through a novel mobile app for fingerspelling. In *Proceedings of the 14th International Conference on Computers Helping People with Special Needs (ICCHP)* (pp. 559–564). Paris, France.

Traxler, C. B. (2000). The Stanford Achievement Test, 9th edition: National norming and performance standards for Deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education*, *5*(4), 337–348.

van der Hulst, H., & Channon, R. (2010). Notation systems. In D. Brentari (Ed.), *Sign languages* (pp. 151–173). New York, NY: Cambridge University Press.

Vilar, D., Stein, D., Huck, M., & Ney, H. (2010). Jane: Open source hierarchical translation, extended with reordering and lexicon models. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and Metrics (MATR)* (pp. 262–270). Uppsala, Sweden.

Wauters, L. (2005). *Reading comprehension in Deaf children: The impact of the mode of acquisition of word meanings* (Unpublished doctoral dissertation). Radboud University Nijmegen.

Wells, J. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of standards and resources for spoken language systems.* Berlin, Germany: De Gruyter Mouton.

Wilbur, R. B. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited* (pp. 215–244). Mahwah, NJ: Erlbaum.

Wilcox, S. (1992). *The phonetics of fingerspelling*. Amsterdam, Netherlands: John Benjamins.

Wolfe, R., Alba, N., Billups, S., Davidson, M. J., Dwyer, C., Jamrozik, D. G., … Young, J. (2006). An improved tool for practicing fingerspelling recognition. In *Proceedings of the International Conference on Technology and Persons with Disabilities* (pp. 17–22). Northridge, CA.

Wolfe, R., Cook, P., McDonald, J. C., & Schnepp, J. (2013). Linguistics as structure in computer animation: Toward a more effective synthesis of brow motion in American Sign Language. *Sign Language & Linguistics*, *14*(1), 179–199.

Xu, W. (2006). *A comparison of Chinese and Taiwan Sign Languages: Towards a new model for sign language comparison* (Unpublished master's thesis). Ohio State University.

Zens, R., & Ney, H. (2008). Improvements in dynamic programming beam search for phrase-based statistical machine translation. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)* (pp. 198–205). Waikiki, HI.

Zeshan, U. (2012). Sprachvergleich: Vielfalt und Einheit von Gebärdensprachen. In H. Eichmann, M. Hansen, & J. Heßmann (Eds.), *Handbuch Deutsche Gebärdensprache* (pp. 311–340). Hamburg, Germany: Signum.

Curriculum Vitae

Personal information

Name	Sarah Rahel Ebling
Date of birth	April 23, 1984
Place of birth	Baden AG, Switzerland

Education

March-Aug. 2015	Research visit, Rochester Institute of Technology, USA (Prof. Dr. Matt Huenerfauth)
Feb. 2015	Research visit, DePaul University Chicago, USA (Prof. Dr. Rosalee Wolfe)
Jan. 2012-Oct. 2014	"Teaching Skills" programme, University of Zurich, Switzerland (completion with certificate)
Since Nov. 2011	PhD student in Computational Linguistics, University of Zurich, Switzerland (advisor: Prof. Dr. Martin Volk)
Oct. 2004-Oct. 2011	Licentiate (Master of Arts), University of Zurich, Switzerland German linguistics and literature, Computational linguistics, English linguistics
April-Dec. 2011	Research visit, School of Computing, Dublin City University, Ireland (Prof. Dr. Andy Way)
Sept. 2008-July 2009	Two-semester Erasmus stay, University of Heidelberg, Germany
2000-2004	Kantonsschule Baden, Switzerland
Aug. 2001-June 2002	Tamalpais High School, Mill Valley, CA, USA

Work experience

Feb.-August 2015	SNSF Doc.Mobility project "Machine Translation and Animation Assessment for Swiss German Sign Language"
Oct. 2013-Oct. 2014	SNSF project "Gaze and Productive Signing in a Corpus of Interactions of Deaf and Hard of Hearing Signers of Swiss German Sign Language (DSGS)" (InterGaze)
Since Nov. 2011	Research assistant, University of Zurich, Switzerland

Jan.-Oct. 2011	Student research assistant, Institute of Computational Linguistics, University of Zurich, Switzerland
Jan.-Feb. 2010	Student research assistant, Heidelberg Institute for Theoretical Studies (HITS) (formerly EML Research), Heidelberg, Germany
Aug. 2008-Dec. 2010	Student research assistant, Project “semtracks”, Heidelberg Center for American Studies, Heidelberg, Germany / University of Zurich, Switzerland
July-Sept. 2006	Journalism internship, Neue Zürcher Zeitung (NZZ), Zurich, Switzerland
Juli 2003-Dec. 2008	Freelance newspaper journalist, Aargauer Zeitung, Baden, Switzerland