

Signed Coreference Resolution

Kayo Yin¹, Kenneth DeHaan², Malihe Alikhani³

¹Language Technologies Institute, Carnegie Mellon University

²Department of American Sign Language, Gallaudet University

³Department of Computer Science, University of Pittsburgh

kayoy@cs.cmu.edu, kenneth.de.haan@gallaudet.edu, malihe@pitt.edu

Abstract

Coreference resolution is key to many natural language processing tasks and yet has been relatively unexplored in Sign Language Processing. In signed languages, space is primarily used to establish reference. Solving coreference resolution for signed languages would not only enable higher-level Sign Language Processing systems, but also enhance our understanding of language in different modalities and of situated references, which are key problems in studying grounded language. In this paper, we: (1) introduce Signed Coreference Resolution (SCR), a new challenge for coreference modeling and Sign Language Processing; (2) collect an annotated corpus of German Sign Language with gold labels for coreference together with an annotation software for the task; (3) explore features of hand gesture, iconicity, and spatial situated properties and move forward to propose a set of linguistically informed heuristics and unsupervised models for the task; (4) put forward several proposals about ways to address the complexities of this challenge effectively¹.

1 Introduction

While signed languages are fully-fledged natural languages with sophisticated grammatical systems that are fully comparable to those of spoken languages (Emmorey, 2001), they are also in a completely different modality with such extreme complexity that has yet to be thoroughly studied and understood. Much of our current language technologies are not effective on signed languages, as natural language processing (NLP) modeling approaches are often based on linguistic theories of spoken languages, and expect either speech or written text as input. This results in technology that may be inaccessible to Deaf people where

signed languages are their primary mean of communication and who strongly prefer using their native language than a spoken language (Padden and Humphries, 1988; Glickman and Hall, 2018), thus it is essential to extend NLP to signed languages. On the other hand, most of the recent research in Sign Language Processing (SLP) mainly focus on the visual component of signed languages and fail to address its linguistic challenges, such as coreference resolution (Yin et al., 2021a).

Coreference resolution is a critical component of natural language understanding and higher-level NLP applications including information extraction, text summarization, and machine translation, yet it is completely unexplored for signed languages. Although coreference relations have been studied in sign linguistics, computational models fall short in this area. Resolving coreference in signed languages presents novel challenges as the meaning of pronominal signs are highly dependent on discourse and spatial context (Cormier et al., 2010, 2013). Tackling this problem will help us gain a better understanding of how grounding is achieved across different types of natural languages and in multimodal communication, and broaden the ability of current NLP systems to handle multiple modalities. In addition, achieving automatic coreference resolution for signed languages will enable technologies for Sign Language Translation or provide educational tools for sign language learners, among many more.

In this paper, we introduce Signed Coreference Resolution (SCR) (Figure 1) as a new challenge for coreference resolution and SLP. We present how coreference is established in signed languages and explore its features of gesture, discourse and spatial grounding for modeling. We then develop a software to annotate signed coreference and release **DGS-Coref**, a German Sign Language (DGS) dataset to evaluate SCR models. We propose a novel architecture based on multigraphs and

¹Our code, data and signed coreference annotation software are publicly available at <https://github.com/kayoyin/scr>.

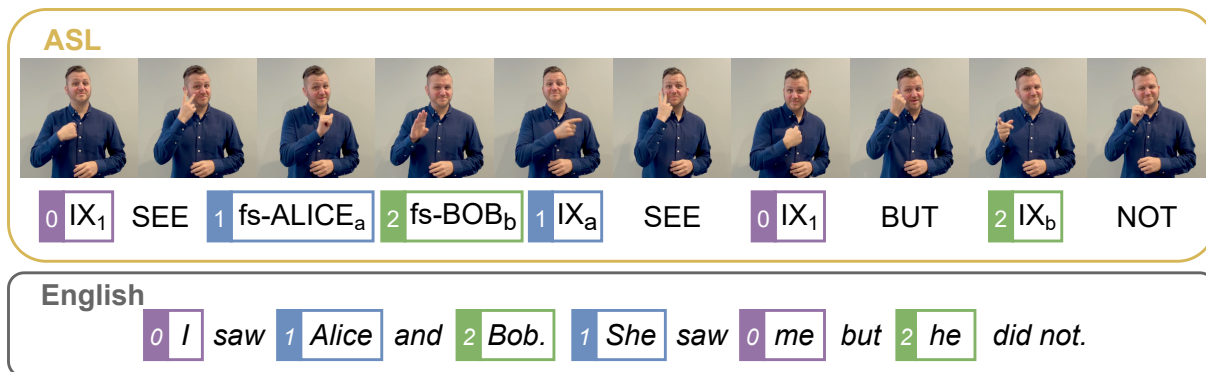


Figure 1: Coreference resolution in American Sign Language (ASL) (video and gloss) and in English. IX_1 indicates an index that points to the signer and IX_a to the location ‘a’. fs-ALICE_a is the name that is fingerspelled at ‘a’. Numbered boxes around ASL glosses and English words correspond to entity labels.

linguistically-informed heuristics to perform unsupervised SCR, which we hope to extend to other signed languages, as all signed languages exhibit similar properties and methods to establish referents in space (McBurney, 2004). Finally, we discuss the complexities and important considerations to take into account for SCR, as well as suggestions for future directions of research. We believe that the development of SCR will provide an important stepping stone to sign language understanding.

2 Related Work

Coreference resolution aims to identify all references to the same entity in discourse and forms a core component of NLP. While automatic coreference resolution has been widely studied for various spoken languages (McCarthy and Lehnert, 1995; Pradhan et al., 2012), no existing work to our knowledge attempts to resolve coreference in signed phrases automatically. We refer to Mitkov (1999) for an overview of the early coreference resolution algorithms, Ng (2010) for the mention-pair model, entity-mention model, and ranking models, and Sukthanker et al. (2020) for a more recent survey of deep-learning based approaches.

2.1 Unsupervised Coreference Resolution

Collecting signed language data is costly, due to the limited availability of qualified signers and annotators, and the complexity of signing videos: 1 hour of a signed video can take up to 100 hours to annotate all manual and non-manual components, compared to 30 hours of annotation for speech (Dreuw and Ney, 2008). Due to the lack of existing annotated signed language data for coreference resolution, we adopt an unsupervised approach.

For unsupervised coreference resolution in spoken languages, earlier works are based on a clustering (Cardie and Wagstaff, 1999; Angheluta et al., 2004) and unsupervised generative models (Haghighi and Klein, 2007; Ng, 2008; Charniak and Elsnar, 2009; Ma et al., 2016). However, these approaches require unannotated training data to learn the model parameters, which are considerably difficult to obtain for the majority of signed languages. Multi-pass sieve systems (Haghighi and Klein, 2009; Raghunathan et al., 2010; Lee et al., 2011, 2013) were popular and effective before the advent of deep learning. However, the sieves used for English cannot be directly applied to signed languages, and it is unclear whether such architecture provides an advantage in our setting while linguistic tools such as POS tags the sieves rely on are not available for signed languages, and the nature of sieves for SCR is unexplored.

Martschat (2013) uses a multigraph-based approach that models a document as a graph, where edges between mentions are established through heuristics. Unlike other graph-based approaches, this method does not need to learn edge weights and therefore remains fully unsupervised. However, it uses constant edge weights which does not account for features with variable strengths. We, instead, build on this approach by proposing novel linguistically-informed heuristics for signed languages, and assign continuous-valued edge weights conditionally to the strength of pair-wise mention features.

2.2 Sign Language Processing

Some of the previously explored SLP tasks include detection (Borg and Camilleri, 2019; Moryossef

et al., 2020), recognition (Imashev et al., 2020; Sincan and Keles, 2020; Cui et al., 2017; Camgöz et al., 2018, 2020b), translation (Camgöz et al., 2018, 2020b; Yin and Read, 2020a,b; Ko et al., 2019; Camgöz et al., 2020a) and production (Stoll et al., 2018, 2020; Saunders et al., 2020a,b; Zelinka and Kanis, 2020; Xiao et al., 2020). However, these efforts often focus mainly on the visual aspect of signed languages without addressing their underlying linguistic structure. Hence, existing SLP models are unable to handle ambiguous pronominal signs, and coreference resolution remains an unaddressed challenge in SLP.

3 Coreference in Signed Languages

Coreference is a core property of natural language (Jackendoff, 2002) and signed language is no exception. Expressed in the visual modality, signed languages use space to maintain discourse coherence and refer back to previously mentioned entities (Liddell, 1980; Kegl, 1987). Moreover, research suggests that the ability to use space to ground referents is innate to humans (Coppola and So, 2006). Therefore, studying coreference in signed languages will give us a better understanding of fundamental phenomena of natural language and help us build tools in various communication systems that are expressed in the visual modality.

3.1 Pronominal Pointing Signs

Sign linguists generally recognize the existence of signs serving a pronominal function in various signed languages (e.g. Van Hoek (1992); Emmorey and Lillo-Martin (1995); Emmorey and Falgier (2004); íc Ciciliani and Wilbur (2006); Cormier et al. (2010)). Referents of pronominal signs are often established in the *signing space*.² The signer can point to the actual location of the referent, such as towards themselves for “I”, towards the addressee for “you”, or towards an entity in the same room for “he, she, they, it”. For entities that are not present, the signer can assign a *locus*³, to the entity, then point at this locus for all mentions of the entity. For example, in Figure 1, the two characters Alice and Bob are introduced by fingerspelling⁴ their names on the left and right side of the signer respectively. To explicitly ground them

²The three-dimensional space in front of the signer used to produce signs

³A particular point in the signing space. Plural *loci*.

⁴Signing a word by spelling out the letters of the word using a manual alphabet.

in the signing space, the signer can also point to the assigned locus after each fingerspelling, although this is not always required. Then, instead of fingerspelling their names at each subsequent mention of one of the characters, the signer can simply point to the locus assigned to the character. Here, the indexing signs serve a similar pronominal function as “she” and “he” in English, and the visual space is heavily exploited to make referencing clear. As a result, the meaning of pronominal pointing signs is not stable and highly depends on its context, and therefore coreference resolution is necessary to identify the antecedent of these signs.

3.2 Complexities of Pointing Signs

There are several complexities in pointing signs to consider during their modeling. For instance, besides a pronominal one, pointing signs may serve other functions as well: as an example, locative pointing signs point to a space to refer to that location instead of a referent (Özyürek et al., 2010), similarly to adverbs “here” and “there” in English, and determiner pointing signs occur in a noun phrase to assign a new locus to an entity (Cormier et al., 2013). Current SLP systems often process solely the local visual features of signs, such as handshape, facial expressions and movement, and therefore cannot disambiguate pointing signs with different meanings and functions that, removed from discourse and spatio-temporal context, have identical visual features.

To compare with spoken languages, while an English pronoun, such as “he”, “she”, “they” carry some meaning on their own, such as the gender or number of the referent, pronominal signs often use the same indexing handshape for personal pronouns, or an open hand with no spaces between fingers for possessive pronouns, regardless of the referents. On the other hand, while the same pronoun “she” can refer to two or more distinct entities at once (for example “My mother never liked Alice, she thought she was up to no good”), a given locus refers to at most one referent at a time. However, the same locus can be reassigned to different entities, and a signed entity may also be assigned one locus to another during a given discourse. Therefore, models must be able to handle long-term dependencies as well as detect and keep track of when there is a change in entity-locus assignment.

Another notable feature of spatial grammar in signed languages is the role of iconicity. Loci

in signed languages can simultaneously have a grammatical (i.e. pronominal) and a logical function. Signed languages observe an iconic semantics where some geometric properties of signs in signing space reflect those in real life, such as the relative positions or sizes of different entities (Schlenker, 2018). Therefore, studying the integration of iconicity and situated referents in signed communication will provide valuable insights in understanding grounded spoken language as well.

While signed languages can help us better understand multimodal communication and linguistic universals in general (Sandler and Lillo-Martin, 2006), some theories of coreference in spoken languages may be extended to signed languages as well. For instance, Steinbach and Onea (2016) extends the classical Discourse Representation Theory (Kamp et al., 2011) to DGS by incorporating the geometrical properties of loci in signing space where discourse referents are grounded. Moreover, Wienholz et al. (2020) finds evidence of the first mention effect in DGS as well (Gernsbacher and Hargreaves, 1988). This suggests that several properties of coreference observed in spoken languages are shared across modalities, which further motivates the development of linguistically-informed SLP models for NLP challenges. We therefore propose to extend the task of coreference resolution to signed languages.

3.3 Signed Coreference Resolution

We formalize the novel challenge of Signed Coreference Resolution (SCR) by decomposing it into two tasks:

Mention Detection Given a video of signing S , we extract all mentions $\{m_1, m_2, \dots, m_N\}$, that is the signs or group of signs in the video that refer to some entity. This task would first require the visual processing of multiple manual and non-manual features in the video to identify each sign, as well as the modeling of long-term dependencies between different signs to deduce mentions. A related existing task is Continuous Sign Language Recognition (CSLR) (Cui et al., 2017; Camgöz et al., 2018, 2020b) that extracts all signed glosses⁵ from a video, though mention detection requires an additional step to group glosses and detect mentions.

There are two possible ways to perform this task:

⁵Glossing refers to the sign-by-sign transcription of signed language.

either mention detection is performed at once during visual processing, where a single pipeline outputs mentions from videos, or CSLR is first performed to extract all glosses, which are then analysed to identify mentions. The advantage of the first method is that it can make full use of all visual features for mention detection and mitigates the bottleneck of an intermediate glossing step. However, SLP research is still at its infancy and CSLR alone is still an ongoing challenge, therefore it may benefit from decomposing the task into several parts. Signed language datasets used for SLP often contain gloss annotations (Cihan Camgöz et al., 2018; Hanke et al., 2020a), therefore it is possible to model mention detection directly on glosses to remove the overhead of visual processing.

Coreference Resolution After obtaining the mentions $\{m_1, m_2, \dots, m_N\}$, we then identify its coreference assignment $C = \{c_1, c_2, \dots, c_N\}$, where each mention m_i is assigned a random variable c_i taking values in the set $\{1, \dots, i\}$. If a set of mentions $\{m_i | i \in \mathcal{I}\}$ all refer to the same entity, then for all $i \in \mathcal{I}, c_i = \min(\mathcal{I})$. For all mentions m_j that do not refer to the same entity as another mention, $c_j = j$.

4 Data

To evaluate SCR models, we develop a small dataset of a signed language with gold coreference labels.

The Public DGS Corpus (Hanke et al., 2020b) is a dataset comprising 50 hours of annotated dialogue between two native signers of DGS. We use this dataset for the following reasons: (i) it is the largest publicly available dataset of a signed language containing gloss annotations at the time, which enables the extraction of enough instances of pronominal pointing to train our models; (ii) it is an open-domain collection of natural signing by 330 native signers, which more closely portray signing in the real-world than other datasets (Yin et al., 2021b); (iii) its annotations include pose estimations, specific glosses for different indexing signs as well as English and German translations, which we use during our modeling.

Although our study is limited to DGS, primarily because of the lack of adequate resources in other signed languages, research suggests all (studied) signed languages use signing space similarly to ground discourse entities and establish pronominal references (McBurney, 2004). Thus, we believe

Task 1 (Video b'1429737', 84) - Example 61
 Video: https://www.sign-lang.uni-hamburg.de/meinedgs/html/1429737_en.html#00053952

English context:
 A: Now I have knee and back pain.
 A: That's why I had to stop.
 A: I was active in the club for over ten years.
 A: Oh well.
 A: I haven't done sports actively here in North Rhine-Westphalia.
 A: I'm working as a sign language teacher.
 A: Back in Berlin I didn't work as a sign language teacher.

English:
 A: When I came here, my partner told me that I would be a great sign language teacher.

English context you highlighted:
[Reset Highlights](#)

English sentence you highlighted:
[Reset Highlights](#)

Glosses context:
 NOW1* I2 KNEE1A* PAIN3 \$GEST-OFF^* LOWER-BACK1E PAIN3
 I1 FINISH1
 OVER-OR-ABOUT1* YEAR1A* ACTIVE1 I1
 \$GEST-OFF^*
 HERE1 NOT1*
 TO-SIGN1A LECTURER1
 PAST-OR-BACK-THEN1* BERLIN1A* \$INDEX1 I1 TO-SIGN1A
 LECTURER1 NOT3A I1*

Glosses:
\$INDEX1 THROUGH2A TO-COME1 \$INDEX1* \$GEST-DECLINE1^ MY1*
 LIFE-PARTNER1 \$INDEX1 TO-RECOMMEND1A* TO-SAY1 TO-MATCH1
 TO-SIGN1A TO-MATCH1

Gloss context you highlighted:
 • BERLIN1A*
 • \$INDEX1

[Reset Highlights](#)

Gloss sentence you highlighted:
[Reset Highlights](#)

How confident are you?

Not at all

Somewhat

Very

Figure 2: Our annotation interface for pronominal indexing signs. For each annotation, the link to the signing video is shown at the top of the page. Annotators are given the previous 7 sentences as the context, with both the English translations (left) and the gloss annotations (right) from the original dataset. The gloss of the sign to be annotated is underlined, and annotators can annotate all glosses shown on the screen that refer to the same entity as the underlined gloss by highlighting them. Annotators can also report their confidence level for each annotation.

that the task we define and the modeling approach we propose are easily generalizable to other signed languages.

4.1 Coreference Annotation

While existing annotated sign language datasets sometimes contain glosses for signs or translations of the signed phrase in a spoken language, none of them contain explicit annotation for coreference. We therefore enhance the gloss annotations from the Public DGS Corpus to construct our dataset.

To do so, we develop an annotation interface for signed languages (Figure), as existing annotation tools for signed languages, especially for targeted tasks such as SCR, are scarce. For each video, our software displays a signed sentence containing a pronominal sign to annotate, accompanied by its English translation. Because coreference can often span several sentences, the previous contextual sentences are also displayed, both in gloss form and in English. The annotator may also play

the video of the phrase being signed along with timestamped gloss annotations. Annotations are submitted by highlighting all glosses shown that refer to the same entity as the underlined gloss. We hired ASL students who were paid 15\$/hour, and our data collection process was approved by our institution's human subject review board. We obtain an inter-annotator agreement of 93.93 in terms of MUC score (Vilain et al., 1995), which suggests high agreement.

4.2 DGS-Coref Dataset

We release the **DGS-Coref** Dataset, a subset of the Public DGS Corpus that has been enhanced with annotations for pronominal indexing coreference. **DGS-Coref** comprises 16 minutes 30 seconds of signing from 3 different conversations featuring 5 different signers. It is composed of 288 signed sentences with 1,457 total glosses, including 95 ⟨I⟩ signs where the signer points towards their chest, 8 ⟨YOU⟩ signs where the signer points to the ad-

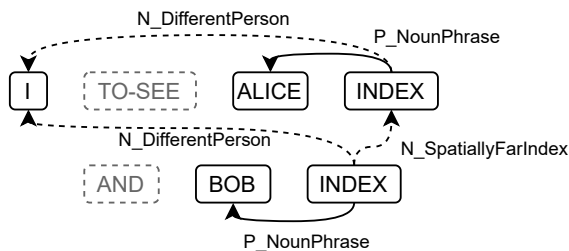


Figure 3: Example of a multigraph constructed from a signed phrase. Solid arrows represent edges with positive weight, dashed arrows represent edges with negative weight. Black glosses are predicted to be mentions and gray glosses are not mentions. The multigraph is constructed by drawing a directed edge for each relation a pair of signs verifies.

dressee, and 93 ⟨INDEX⟩ signs for other pointing signs.⁶

5 Model

In this initial study of SCR, we use DGS glosses and spatial features extracted from pose estimations of the signed phrases to remove the overhead of visual processing and model the linguistic aspect of the task while adequate sign language recognition resources are lacking. We jointly model mention detection and coreference resolution on glosses.

5.1 Unsupervised Continuous Multigraph

The backbone of our approach is based on the unsupervised multigraph coreference model [Martschat \(2013\)](#). The advantage of this model is that it achieves competitive performance in unsupervised coreference resolution in English, while not requiring large unannotated data to tune parameters, which is not always readily available in signed languages. Moreover, its architecture is flexible in allowing the modeling of features with various importance, which is especially adapted to the continuous nature of signing space.

We model the document as a directed labeled weighted multigraph $D = (R, V, A, w)$. Two mentions $m, n \in V$ are two nodes of the graph, and have a directed edge $e = (m, n, r) \in A$ with weight $w(e)$ and label $r \in R$ if m precedes n and the relation $r(m, n)$ holds true (Figure 3). Then, clustering is applied to the resulting multigraph to obtain the entity groups contained in the document.

⁶By convention, we will refer to sign glosses using all capitals

5.2 Relations

First, we define a set of relations that either suggests coreference between two candidate mentions, or provides constraints against possible coreference candidates. Previously explored coreference relations for spoken languages often rely on lexical heuristics and linguistic features such as syntactic dependencies, part-of-speech tags, or morphology. However, such features are currently not available for signed languages due to the lack of core NLP tools to provide them and the recency of linguistic studies on signed languages to develop such tools. Moreover, coreference is inherently expressed differently between spoken and signed languages, which motivates us to design a new set of indicators and constraints for coreference.

First, we propose the following heuristics as positive relations that indicate of coreference:

(1) **P_IAndI** The two signs are produced by the same signer and point to the signer’s chest.

(2) **P_YouAndYou** The two signs are produced by the same signer and point away from the signer’s body towards the addressee.

(3) **P_IAndYou** The two signs are produced by different signers, one points to the signer’s chest and the other points away from the signer’s body towards the addressee.

(4) **P_TemporallyCloseIndex** The two signs are indexing signs produced by the same signer and have less than 10 signs between them.

(5) **P_NounPhrase** If an indexing sign has no other indexing signs within the previous 10 signs, it is coreferent to the temporally closest previous sign, that is not a verb, produced by the same signer.

(6) **P_SpatiallyCloseIndex** The two signs are indexing signs produced by the same signer and the Euclidean distance between the two locations of production is less than 50 pixels.

We also add constraints to coreference as *negative* relations:

(7) **N_IAndI** The two signs are produced by different signers and point to the respective signer’s chest.

(8) **N_YouAndYou** The two signs are produced by different signers and point to the respective addressee.

(9) **N_IAndYou** The two signs are produced by the same signer, one points to the signer’s chest and the other points away from the signer’s body towards the addressee.

(10) **N_DifferentPerson** One sign either points

	MUC			B ³			CEAF _e			Mean
	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	F1	F1
all	67.94	59.55	63.47	61.68	53.03	57.03	22.18	52.08	31.11	50.54
⟨I⟩/⟨YOU⟩	96.7	100	98.32	87.75	100	93.48	95.4	76.32	84.8	92.2
⟨INDEX⟩	64.7	15.49	24.99	88.13	17.55	29.28	18.35	47.32	26.44	26.90

Table 1: Results of our unsupervised continuous multigraph on DGS-Coref. With a coreference score of 50.54, our model provides a strong baseline for SCR, while ⟨INDEX⟩ signs show the most room for improvement.

towards the signer’s chest or towards the addressee, and the other points to a third location.

(11) N_SpatiallyFarIndex The two signs are indexing signs produced by the same signer and the Euclidean distance between the two loci of production is greater than 100 pixels.

5.3 Weight Assignment

For all negative relations, we assign the weight $w(e) = -\infty$ as they are hard constraints for coreference. For binary positive relations (relations 1-3), we assign a fixed weight $w(e) = 0.5$.

Because spoken languages are discrete in nature, it is reasonable that previous work models coreference with fixed weights. However, in signed languages, referents are grounded in continuous time and space, and we hypothesize that the temporal or spatial proximity of signs are strong signals for coreference. Therefore, we introduce a novel continuous weighting system to our model.

For **(4) P_TemporallyCloseIndex** and **(5) P_NounPhrase**, if the signs m and n have $k < 10$ signs between them, the assigned weight is $w(e) = (10 - k)/20$.

For **(6) P_SpatiallyCloseIndex**, if the Euclidean distance between the signs m and n is $k < 50$, the assigned weight is $w(k) = (50 - k)/50$. We assign stronger weights to spatially close indexing signs than temporally close ones, based on the hypothesis that referencing in signed languages are mostly grounded in space.

5.4 Clustering

We apply 1-nearest-neighbor clustering on the obtained multigraph to identify coreferent signs: for every sign n , its candidate antecedents are all signs m such that there exists at least one edge $e = (m, n, r) \in A$, and the sum of edge weights between m and n is strictly positive. n is a mention if it has at least one candidate. If n is a mention, the antecedent of n is the candidate whose sum of edge weights with n is maximal. Ties for antecedents

are broken by selecting the closest sign temporally.

6 Results

In this section, we discuss the strengths and limitations of our approach. As SCR is a new challenge with no existing baseline, our proposed unsupervised model presents a strong baseline for subsequent works.

6.1 Quantitative Evaluation

We evaluate our system on commonly used metrics for coreference resolution in spoken languages: MUC (Vilain et al., 1995), B³ (Bagga and Baldwin, 1998), and CEAF_e (Luo, 2005). We use the official CoNLL shared task scorer⁷.

Table 1 shows the full results of our model. We achieve a mean F1 of 50.54 across all indexing signs. Overall, we achieve a high mean F1 of 92.2 on ⟨I⟩ and ⟨YOU⟩ signs, which is expected as they contain low ambiguity in meaning. On ⟨INDEX⟩ signs, where the model must keep track of spatial coherence in discourse and resolve different loci, we achieve 26.9 mean F1, which shows there is still much room for improvement to disambiguate third-person indexing signs.

⟨INDEX⟩ signs obtain the lowest F1 on the MUC metric, which focuses on the links between pairs of mentions, therefore is especially penalized when there are either extra or missing links in the prediction. On the other hand, ⟨INDEX⟩ signs obtain the highest F1 on the B³ metric, which is a mention-based metric and scores are computed based on individual mentions rather than links. This can lead to the mention identification effect (Moosavi and Strube, 2016), where the metric unreliably rewards mentions that are correctly identified, but linked to the wrong entity, and suggests that our model may be able to detect mentions accurately but is weaker at finding the correct links.

⁷<https://github.com/conll/reference-coreference-scorers>



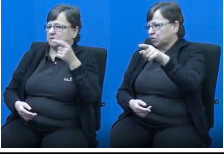
	Relation	Video
<p>TO-SEE YOU GOOD YOU</p> <p><i>I think you could do a good job there.</i></p>	(2) P_YouAndYou	
<p>GEST-DECLINE1 I CAN NOT TO-SAY TO-HOLD-ON I</p> <p><i>I can't keep that promise.</i></p>	(1) P.IAndI, (3) P.IAndYou	
<p>STUTTGART NUM-1 NAME INDEX NUM-1 FREIBURG</p> <p><i>Once we were in Stuttgart, once in Ingolstadt and once in Freiburg.</i></p>	(5) P_NounPhrase	
<p>WITH TRIP INDEX SHIP INDEX</p> <p><i>We went there with an excursion boat.</i></p>	(4) P_TemporallyCloseIndex (6) P_SpatiallyCloseIndex	
<p>I TO-LEARN INDEX HAMBURG INDEX</p> <p><i>I learned it in Hamburg.</i></p>	(4) P_TemporallyCloseIndex (6) P_SpatiallyCloseIndex	

Table 2: Qualitative analysis of model outputs, with relations that were applied for the prediction and the video frames of the glosses in bold. **Bold** glosses are mentioned predicted by our model as coreferent. Underlined glosses are ground-truth coreferent mentions. English translations are provided in italics.

6.2 Qualitative Analysis

To go beyond the limitations of automatic coreference metrics and investigate how our system handles various phenomena in pronominal indexing signs, we perform a qualitative analysis of our model outputs. In Table 2, we give examples of our model outputs and the gold annotations. The first example shows how most coreference relations with ⟨I⟩ and ⟨YOU⟩ are effectively handled by our system. The second example demonstrates how the model can detect the introduction of a new referent to the discourse and signing space. In the third example, the model successfully resolves the two indexing signs as coreferent, due to their temporal and spatial proximity.

In the last example, the model fails to identify “Hamburg” being introduced as a new referent. Instead, it resolves the second ⟨INDEX⟩ to the first, as relations (4) and (6) give stronger

weights to the multi-edge between the two indexing signs than the relation (5) does to the edge between ⟨HAMBURG⟩ and ⟨INDEX⟩. In general, the main weakness of our model is choosing correctly between an antecedent candidate that is a spatially close indexing sign and another candidate that marks the introduction of a new referent. To overcome this challenge, we believe that a more sophisticated system to model the deeper meaning of the signed phrase is needed.

6.3 Discussion

We now discuss phenomena that are beyond the scope of this initial study, but that are important challenges and considerations to take for future efforts in SCR.

Naturally, signers may reassign the locus to a new referent, which our current approach does not explicitly address and can only capture this if the locus is reassigned after an extended period of not

being used, which is not always the case. Future approaches need to be able to detect when a change of referent for a locus occurs.

As discussed in §3, not all indexing signs are pronominal either, some may serve a locative function where it is not necessarily coreferent with another sign in discourse, but is used to refer to a physical location in space. Future work should therefore be able to distinguish the different functions of indexing signs.

Finally, the partitioning of signing space is dynamic (Steinbach and Onea, 2016). For example, when there are only two referents established, the locus assigned to each can be relatively large without causing ambiguity, such as the first referent being assigned the right half, and the second the left half of the signing space. As more referents are introduced, the signing space is partitioned into smaller loci. Therefore, what constitutes two indexing signs that are “spatially close enough” to be pointing to the same locus depends on the evolution of the discourse, whereas our approach maintains the same heuristic on spatial relations throughout discourse.

7 Conclusions and Future Work

We present a new challenge for automatically resolving and evaluating coreference in signed languages. We also release the first dataset in German Sign Language with gold labels for coreference resolution, as well as a web interface to annotate coreference in signed languages. Finally, we propose a novel model to perform unsupervised coreference resolution that relies on a multigraph-based architecture with new, linguistically-informed heuristics which provides a strong baseline for this task.

Our paper performs coreference resolution on glosses to remove the overhead of visual processing and focus on the purely linguistic aspect of signed coreference. Future work involves modeling approaches that process signing videos directly that may more closely reflect real-world applications. We also leave for future work the resolution of non-indexing signs that also may serve a pronominal function, such as body shift and facial markers. Our work can additionally be extended to studying other types of ambiguous signs, such as directional verbs where the subject and/or object are not explicitly signed but grounded in space.

This task also provides the opportunity to explore ways studying SCR can benefit spoken lan-

guage understanding, particularly multimodal communication where meaning in spoken languages can also be conveyed through the visual modality, such as co-speech indexing gestures. We also hope that future efforts towards SCR and SLP in general, through close collaboration with signing communities, result in assistive technology that can help deaf students in education, research, and everyday communication in their preferred language.

Acknowledgements

We would like to thank Sushruti Bansod and Kate Atwell for help with the annotation of the SCR dataset. We also thank Julie Hochgesang, Alex Shypula, Marc Schulder, Richard Boyce and Amit Moryossef for their helpful advice and feedback. This project was supported by the University of Pittsburgh Momentum fund for research towards reducing language obstacles that Deaf students face when developing scientific competencies.

References

- Roxana Angheluta, Patrick Jeuniaux, Rudradeb Mitra, and Marie-Francine Moens. 2004. Clustering algorithms for noun phrase coreference resolution. *Proceedings of the 7es Journes internationales d’Analyse statistique des Donnes Textuelles*, pages 60–70.
- Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In *The first international conference on language resources and evaluation workshop on linguistics coreference*, volume 1, pages 563–566. Citeseer.
- Mark Borg and Kenneth P Camilleri. 2019. Sign language detection “in the wild” with recurrent neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1637–1641. IEEE.
- Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.
- Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020a. Multi-channel transformers for multi-articulatory sign language translation. In *European Conference on Computer Vision*, pages 301–319.
- Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020b. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10033.

- Claire Cardie and Kiri Wagstaff. 1999. [Noun phrase coreference as clustering](#). In *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.
- Eugene Charniak and Micha Elsner. 2009. [EM works for pronoun anaphora resolution](#). In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 148–156, Athens, Greece. Association for Computational Linguistics.
- Tamara Aliba ić Ciciliani and Ronnie B Wilbur. 2006. Pronominal system in croatian sign language. *Sign Language & Linguistics*, 9(1-2):95–132.
- Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.
- Marie Coppola and Wing Chee So. 2006. The seeds of spatial grammar: Spatial modulation and coreference in homesigning and hearing adults. In *BU-CLD 30: Proceedings of the 30th Annual Boston University Conference on Language Development*, volume 30, pages 119–130.
- Kearsy Cormier, Adam Schembri, and Bencie Woll. 2010. Diversity across sign languages and spoken languages: Implications for language universals. *Lingua*, 120(12):2664–2667.
- Kearsy Cormier, Adam Schembri, and Bencie Woll. 2013. Pronouns and pointing in sign languages. *Lingua*, 137:230–247.
- Runpeng Cui, Hu Liu, and Changshui Zhang. 2017. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7361–7369.
- Philippe Dreuw and Hermann Ney. 2008. Towards automatic sign language annotation for the elan tool. In *Workshop Programme*, volume 50.
- Karen Emmorey. 2001. *Language, cognition, and the brain: Insights from sign language research*. Psychology Press.
- Karen Emmorey and Brenda Falgier. 2004. Conceptual locations and pronominal reference in american sign language. *Journal of Psycholinguistic Research*, 33(4):321–331.
- Karen Emmorey and Diane Lillo-Martin. 1995. Processing spatial anaphora: Referent reactivation with overt and null pronouns in american sign language. *Language and Cognitive Processes*, 10(6):631–653.
- Morton Ann Gernsbacher and David J Hargreaves. 1988. Accessing sentence participants: The advantage of first mention. *Journal of memory and language*, 27(6):699–717.
- Neil S Glickman and Wyatte C Hall. 2018. *Language deprivation and deaf mental health*. Routledge.
- Aria Haghighi and Dan Klein. 2007. [Unsupervised coreference resolution in a nonparametric Bayesian model](#). In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 848–855, Prague, Czech Republic. Association for Computational Linguistics.
- Aria Haghighi and Dan Klein. 2009. [Simple coreference resolution with rich syntactic and semantic features](#). In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1152–1161, Singapore. Association for Computational Linguistics.
- Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020a. [Extending the Public DGS Corpus in size and depth](#). In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).
- Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020b. [Extending the Public DGS Corpus in size and depth](#). In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).
- Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman, and Anara Sandygulova. 2020. A dataset for linguistic understanding, visual evaluation, and recognition of sign languages: The k-rsl. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 631–640.
- Ray Jackendoff. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, USA.
- Hans Kamp, Josef Van Genabith, and Uwe Reyle. 2011. Discourse representation theory. In *Handbook of philosophical logic*, pages 125–394. Springer.
- Judy Kegl. 1987. Coreference relations in american sign language. *Studies in the Acquisition of Anaphora*, pages 135–170.
- Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural sign language translation based on human keypoint estimation. *Applied Sciences*, 9(13):2683.
- Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2013. [Deterministic coreference resolution based on entity-centric, precision-ranked rules](#). *Computational Linguistics*, 39(4):885–916.

- Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. [Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task](#). In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 28–34, Portland, Oregon, USA. Association for Computational Linguistics.
- Scott K Liddell. 1980. *American sign language syntax*. De Gruyter Mouton.
- Xiaoqiang Luo. 2005. [On coreference resolution performance metrics](#). In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 25–32, Vancouver, British Columbia, Canada. Association for Computational Linguistics.
- Xuezhe Ma, Zhengzhong Liu, and Eduard Hovy. 2016. [Unsupervised ranking model for entity coreference resolution](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1012–1018, San Diego, California. Association for Computational Linguistics.
- Sebastian Martschat. 2013. [Multigraph clustering for unsupervised coreference resolution](#). In *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*, pages 81–88, Sofia, Bulgaria. Association for Computational Linguistics.
- Susan Lloyd McBurney. 2004. *Referential morphology in signed languages*. University of Washington.
- Joseph F McCarthy and Wendy G Lehnert. 1995. Using decision trees for coreference resolution. *IJCAI*.
- Ruslan Mitkov. 1999. *Anaphora resolution: the state of the art*. Citeseer.
- Nafise Sadat Moosavi and Michael Strube. 2016. [Which coreference evaluation metric do you trust? a proposal for a link-based entity aware metric](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 632–642, Berlin, Germany. Association for Computational Linguistics.
- Amit Moryossef, Ioannis Tsochantaridis, Roei Aharoni, Sarah Ebling, and Srini Narayanan. 2020. Real-time sign language detection using human pose estimation. In *European Conference on Computer Vision*, pages 237–248. Springer.
- Vincent Ng. 2008. [Unsupervised models for coreference resolution](#). In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 640–649, Honolulu, Hawaii. Association for Computational Linguistics.
- Vincent Ng. 2010. [Supervised noun phrase coreference research: The first fifteen years](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411, Uppsala, Sweden. Association for Computational Linguistics.
- Aslı Özyürek, Inge Zwisserlood, and Pamela Perniss. 2010. Locative expressions in signed languages: A view from turkish sign language (tid).
- Carol A Padden and Tom Humphries. 1988. *Deaf in America*. Harvard University Press.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. [CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes](#). In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40, Jeju Island, Korea. Association for Computational Linguistics.
- Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, and Christopher Manning. 2010. [A multi-pass sieve for coreference resolution](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 492–501, Cambridge, MA. Association for Computational Linguistics.
- Wendy Sandler and Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge University Press.
- Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. 2020a. Everybody sign now: Translating spoken language to photo realistic sign language video. *arXiv preprint arXiv:2011.09846*.
- Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. 2020b. Progressive transformers for end-to-end sign language production. In *European Conference on Computer Vision*, pages 687–705.
- Philippe Schlenker. 2018. Visible meaning: Sign language and the foundations of semantics. *Theoretical Linguistics*, 44(3-4):123–208.
- Ozge Mercanoglu Sincan and Hacer Yalim Keles. 2020. Auts!: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355.
- Markus Steinbach and Edgar Onea. 2016. A drt analysis of discourse referents and anaphora resolution in sign language. *Journal of Semantics*, 33(3):409–448.
- Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2018. Sign language production using neural machine translation and generative adversarial networks. In *Proceedings of the 29th British Machine Vision Conference (BMVC 2018)*. British Machine Vision Association.

- Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2020. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, pages 1–18.
- Rhea Sukthanker, Soujanya Poria, Erik Cambria, and Ramkumar Thirunavukarasu. 2020. Anaphora and coreference resolution: A review. *Information Fusion*, 59:139–162.
- Karen Van Hoek. 1992. Conceptual spaces and pronominal reference in american sign language. *Nordic Journal of Linguistics*, 15(2):183–199.
- Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. [A model-theoretic coreference scoring scheme](#). In *Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995*.
- Anne Wienholz, Derya Nuhbalaoglu, Nivedita Mani, Annika Herrmann, Edgar Onea, and Markus Steinbach. 2020. Processing pronominal pointing signs in german sign language: Neurophysiological evidence for the first mention effect.
- Qinkun Xiao, Minying Qin, and Yuting Yin. 2020. Skeleton-based chinese sign language recognition and generation for bidirectional communication between deaf and hearing people. *Neural Networks*, 125:41–55.
- Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021a. [Including signed languages in natural language processing](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360, Online. Association for Computational Linguistics.
- Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021b. [Including signed languages in natural language processing](#). In *Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, Virtual.
- Kayo Yin and Jesse Read. 2020a. Attention is all you sign: sign language translation with transformers. In *Sign Language Recognition, Translation and Production (SLRTP) Workshop-Extended Abstracts*, volume 4.
- Kayo Yin and Jesse Read. 2020b. [Better sign language translation with STMC-transformer](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5975–5989, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Jan Zelinka and Jakub Kanis. 2020. Neural sign language synthesis: Words are our glosses. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3395–3403.