Carnegie Mellon University
Language Technologies Institute

Natural Language Processing for Signed Languages

Kayo Yin

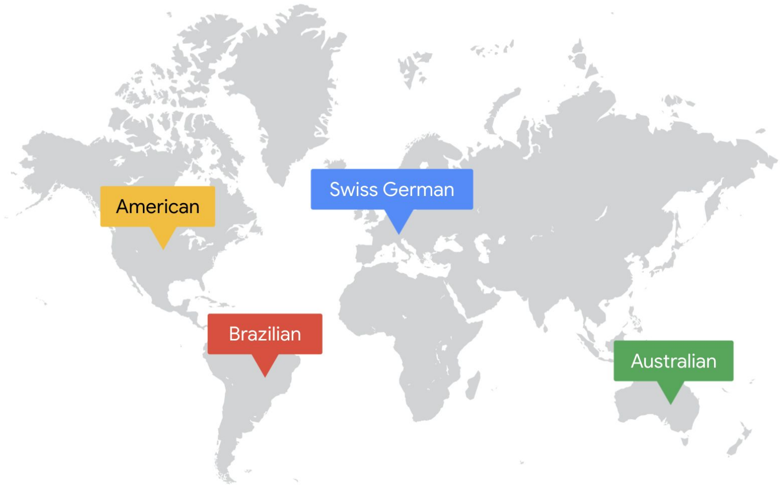
DeepMind SL Reading Group

November 5 2021

Signed Languages

- Fully-fledged natural languages
- Expressed through various cues
- Independent of spoken languages

Signed Languages

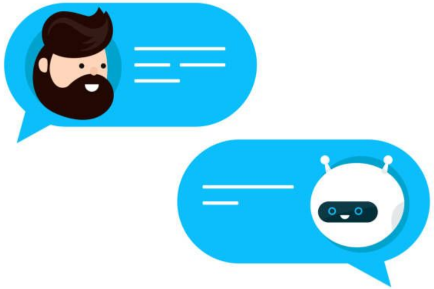


- 200 signed languages
- ~70m deaf people

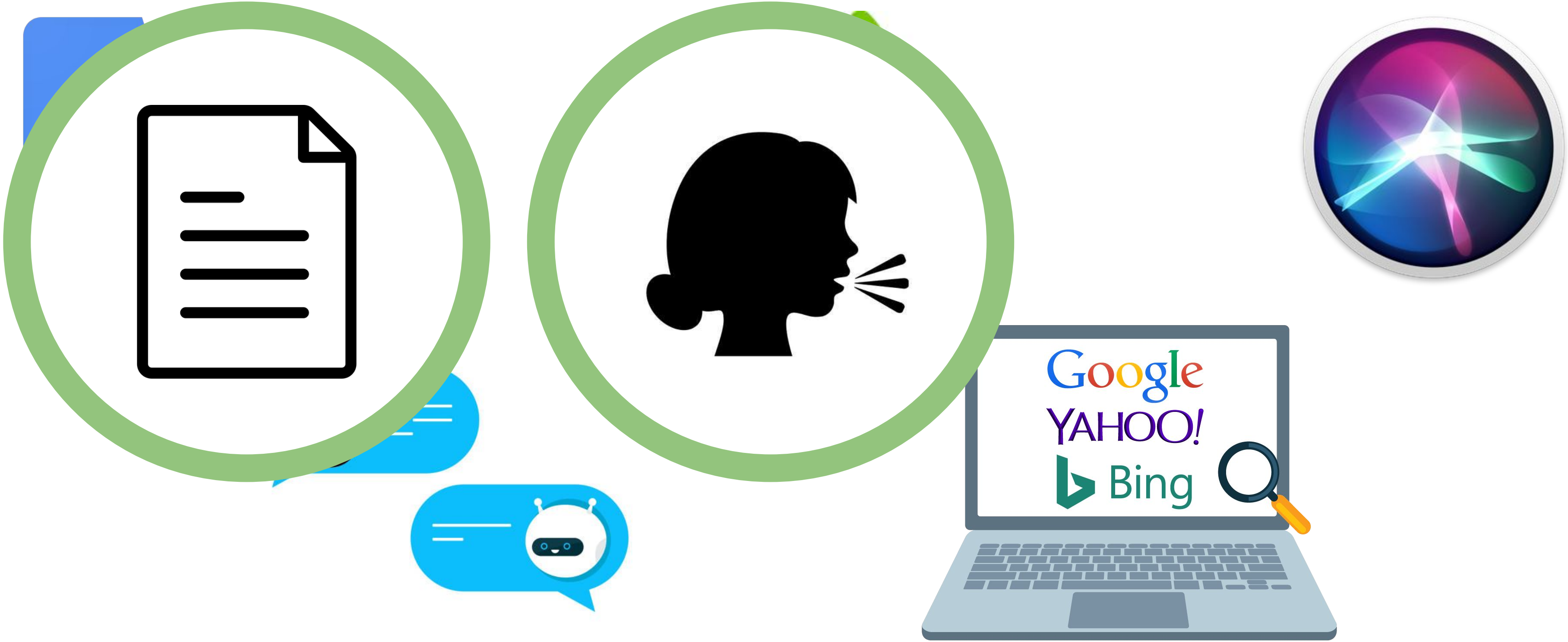
Signed Languages

- Primary and preferred means of communication for Deaf communities

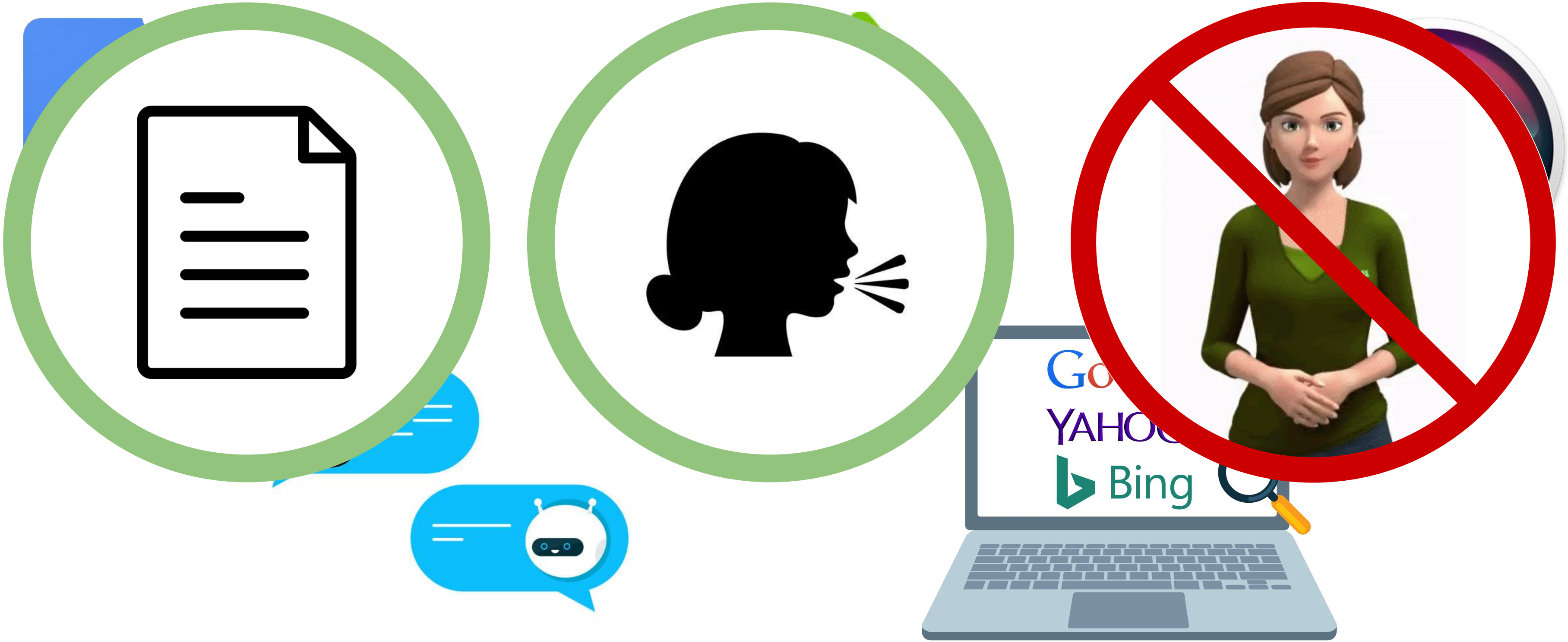
Who Benefits from Natural Language Processing?



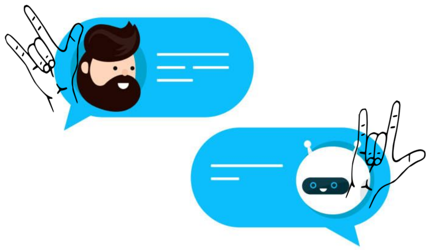
Who Benefits from Natural Language Processing?



Who Benefits from Natural Language Processing?

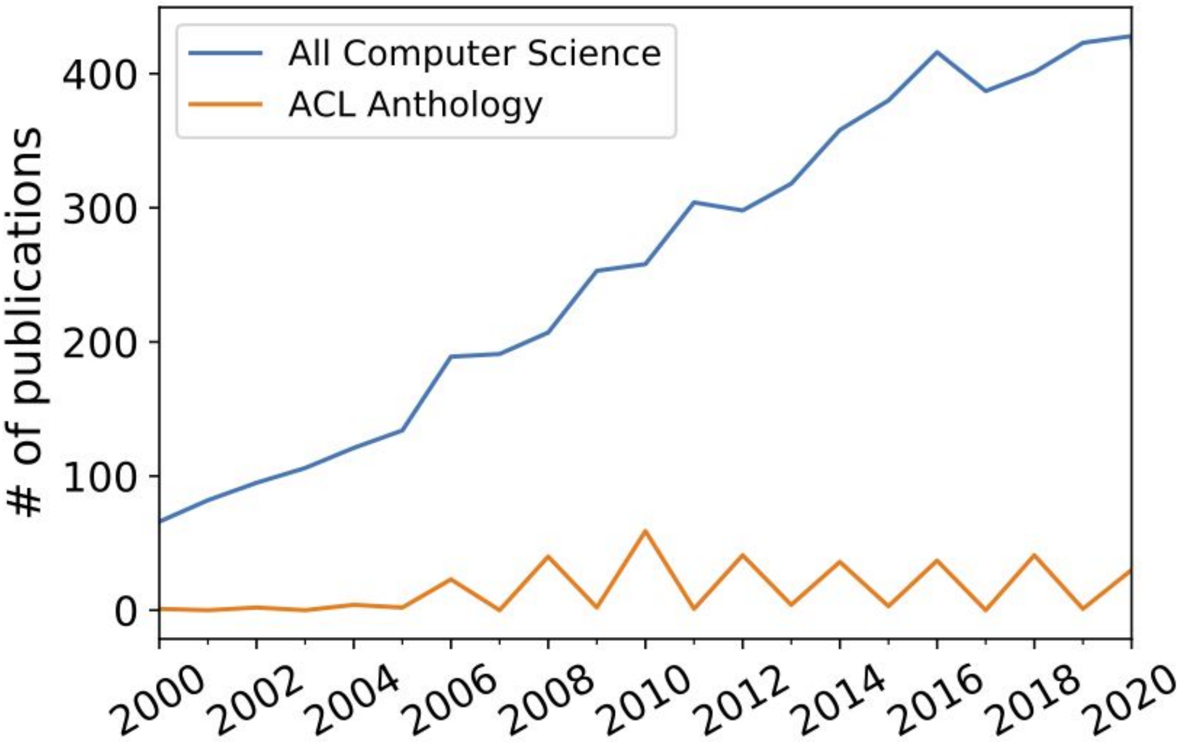


Who Benefits from Natural Language Processing?

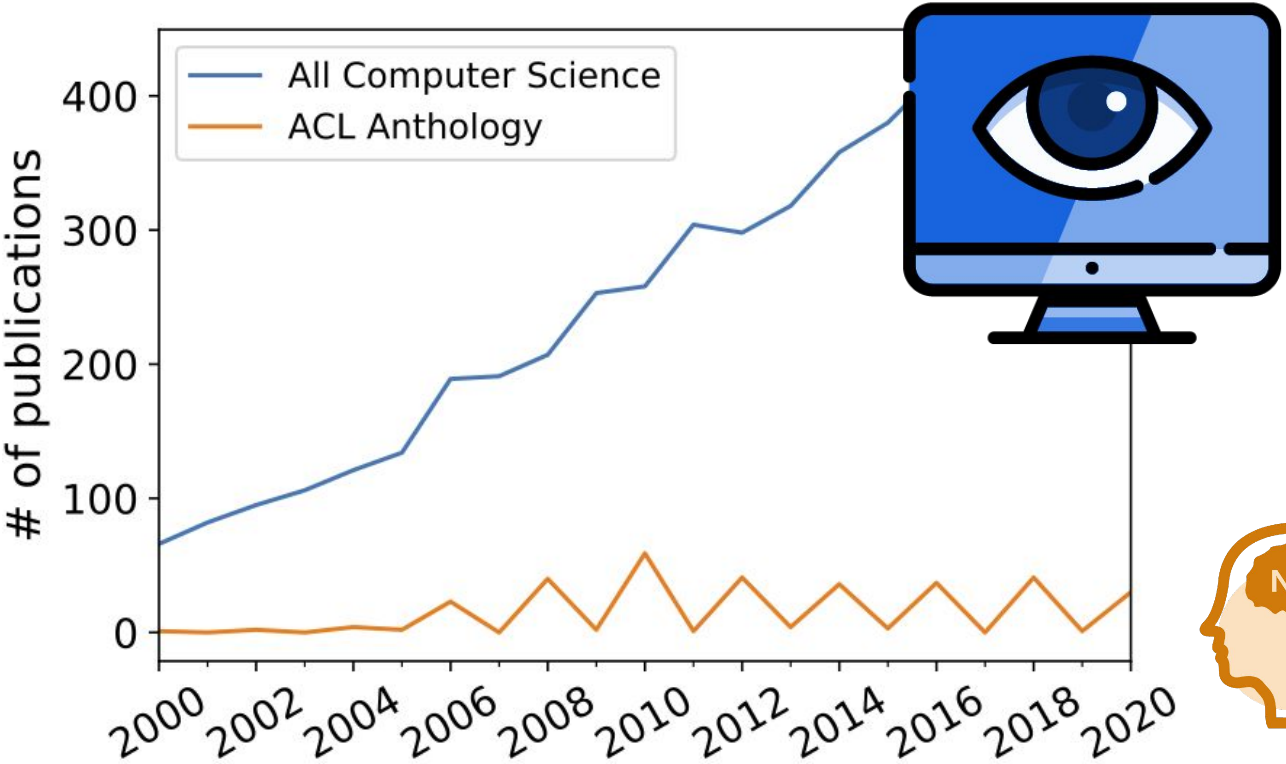


Let’s allow everyone to benefit from technology using their preferred language!

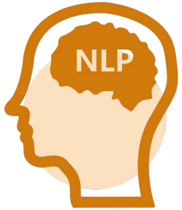
Who is Working on Sign Language Processing?



Who is Working on Sign Language Processing?

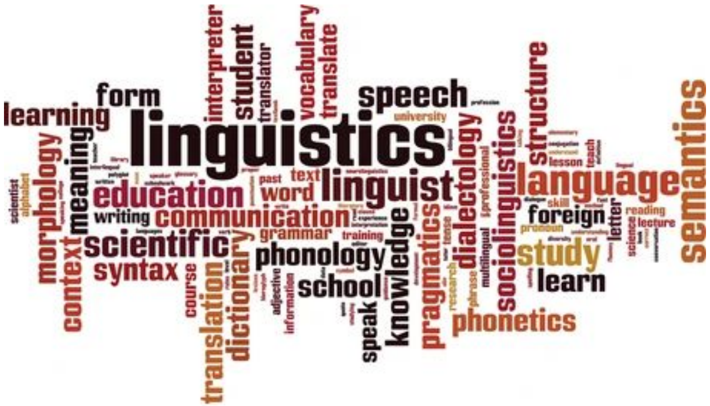


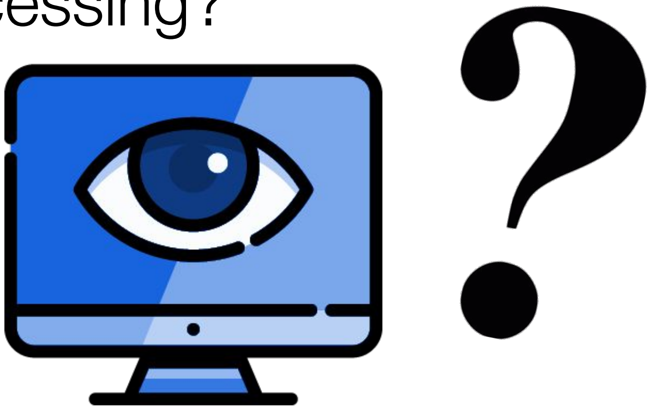
Mostly computer vision



Little NLP involvement

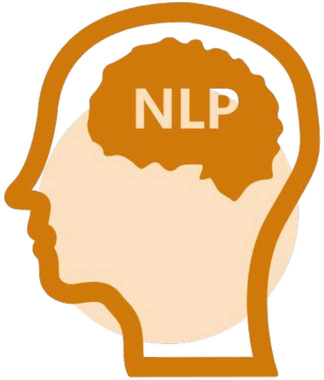
Who is Working on Sign Language Processing?

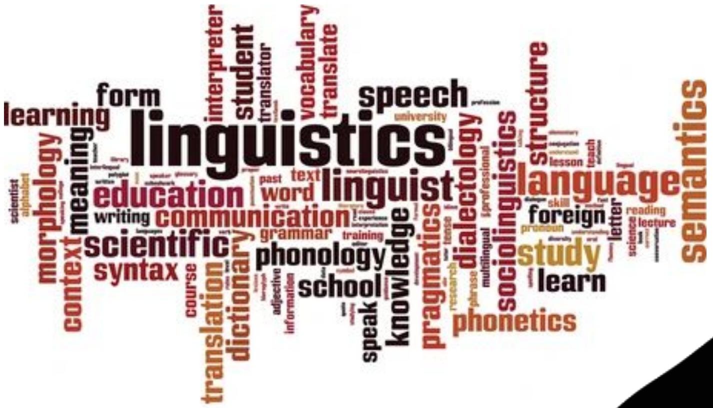


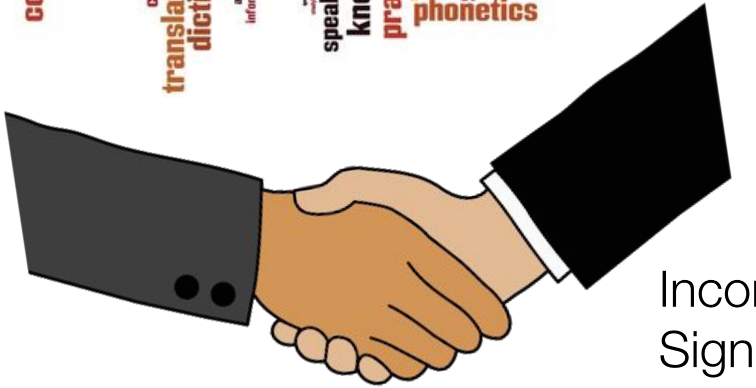


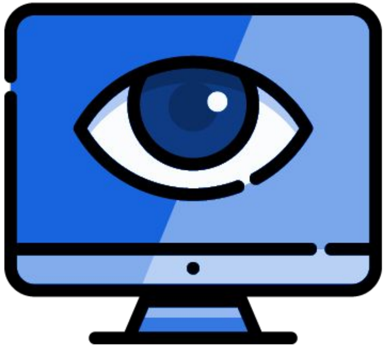
Current models ignore the linguistic structure of signed languages

Who is Working on Sign Language Processing?





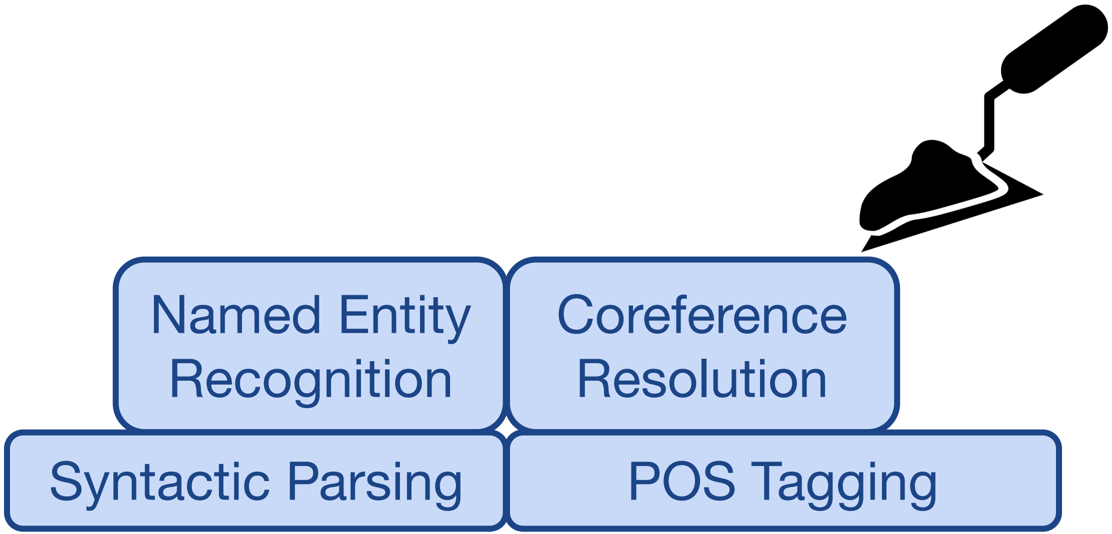




Incorporate linguistic insight into Sign Language Processing

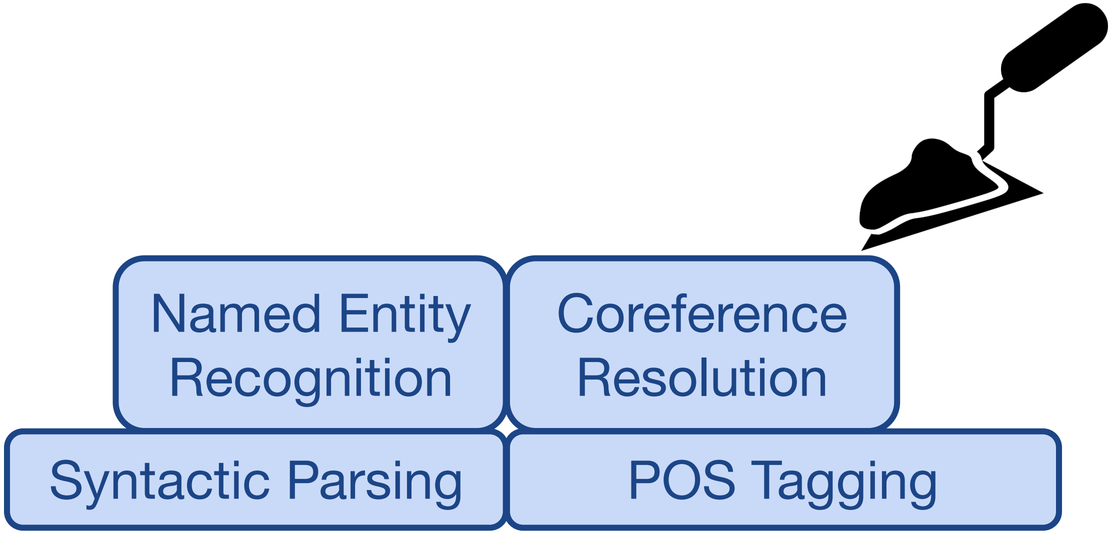
Extending NLP to Signed Languages

- Both spoken and signed languages express the grammar of natural languages

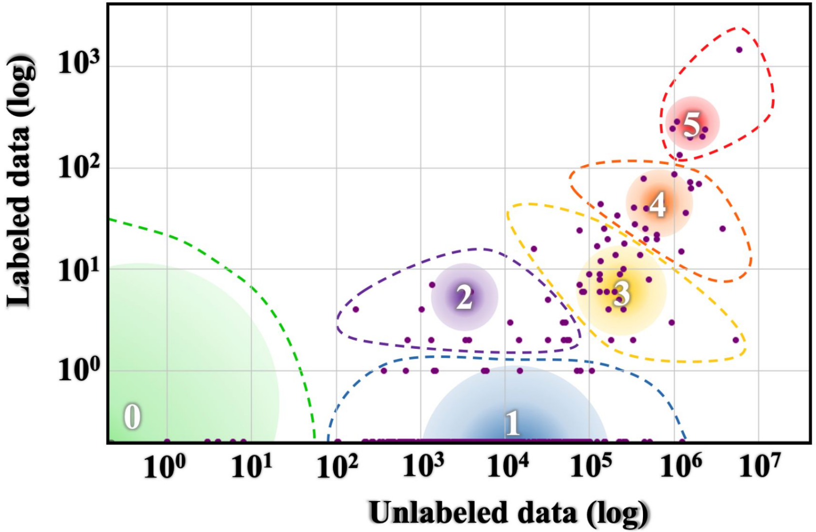


Extending NLP to Signed Languages

- Both spoken and signed languages express the grammar of natural languages
- Extend core NLP tools to signed languages



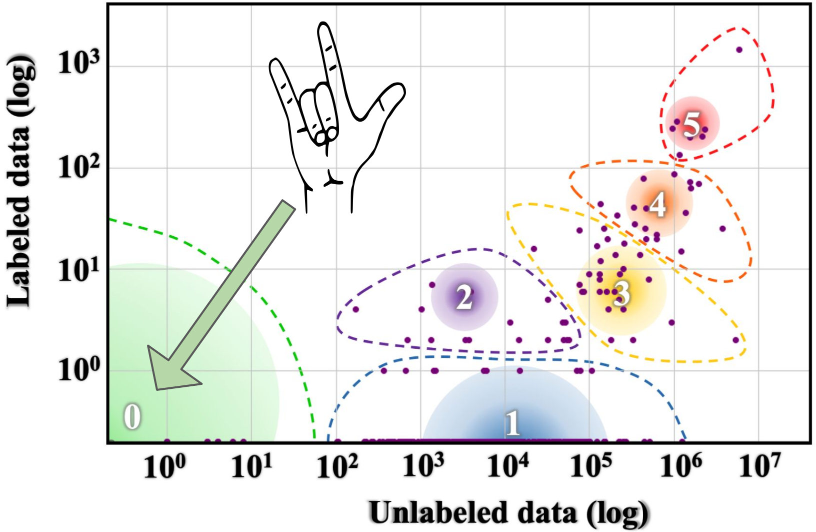
Challenges: Data Scarcity



Taxonomy of language resources
(Joshi et al., [2020](#))

- Need large, realistic datasets

Challenges: Data Scarcity



Taxonomy of language resources
(Joshi et al., 2020)

- Need large, realistic datasets
- All signed languages are extremely low-resource

Challenges: Data Scarcity

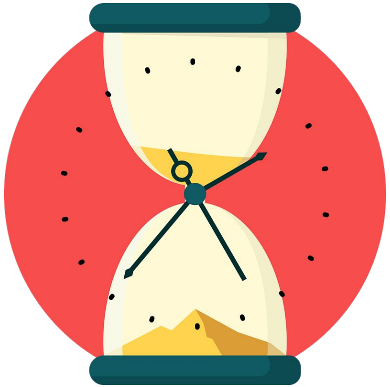
- Difficult to recruit and record signers for data collection

Challenges: Data Scarcity

- Difficult to recruit and record signers for data collection
- Finding / training annotators is challenging

Challenges: Data Scarcity





- Difficult to recruit and record signers for data collection
- Finding / training annotators is challenging
- 1 minute of labelled data requires 600 minutes of data collection

Challenges: Spatial Dependencies

- Grounding in signing space

Challenges: Spatial Dependencies

- Grounding in signing space
- We need to model the spatial discourse

Natural Language Processing for Signed Languages

In this talk, we explore:

- **Data augmentation** for Sign Language Translation

Natural Language Processing for Signed Languages

In this talk, we explore:

- **Data augmentation** for Sign Language Translation
- **Coreference resolution** for pronominal indexing signs

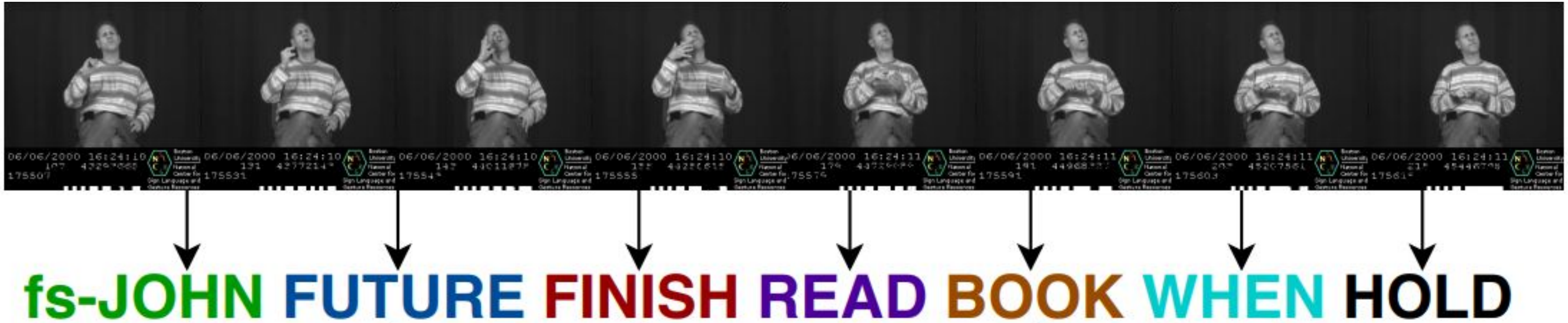
Data Augmentation for Sign Language Gloss Translation

Amit Moryossef*, Kayo Yin*, Graham Neubig, Yoav Goldberg

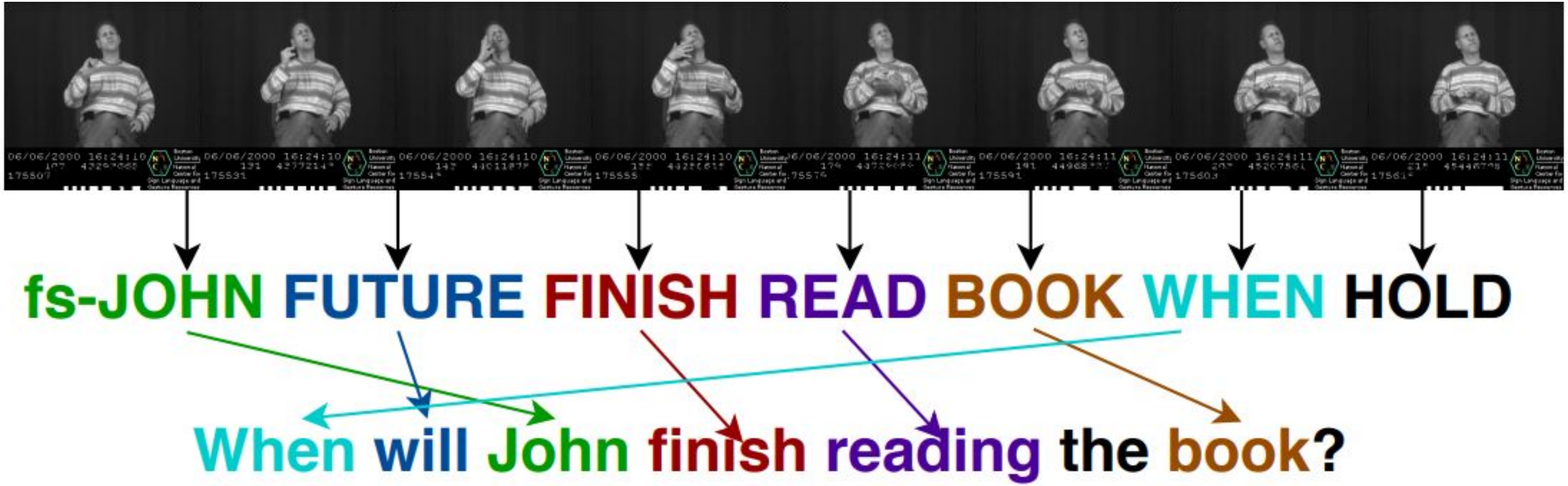
(MTSummit21 AT4SSL Workshop)

*Equal contribution

Sign Language Translation



Sign Language Translation



Overcoming Data Scarcity

- Gloss-to-text translation = **extremely low resource** MT

Overcoming Data Scarcity

- Gloss-to-text translation = **extremely low resource** MT
- How is the relationship between a signed and spoken language **different** from two spoken languages?

Overcoming Data Scarcity

- Gloss-to-text translation = **extremely low resource** MT
- How is the relationship between a signed and spoken language **different** from two spoken languages?
- Can we improve gloss-to-text translation using **pseudo-parallel data**?

Signed vs. Spoken Languages

- Lexical similarity

- Syntactic similarity

Signed vs. Spoken Languages

- Lexical similarity

$$o_w = \frac{|T_1 \cap T_2|}{|T_1| + |T_2|}$$

- Syntactic similarity

Signed vs. Spoken Languages

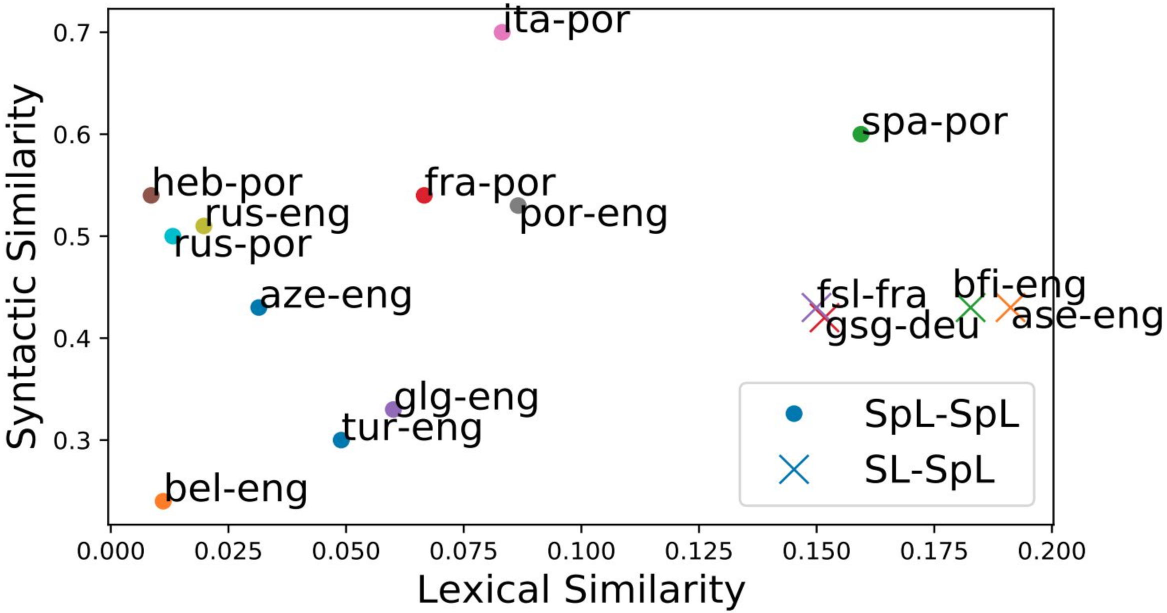
- Lexical similarity

$$o_w = \frac{|T_1 \cap T_2|}{|T_1| + |T_2|}$$

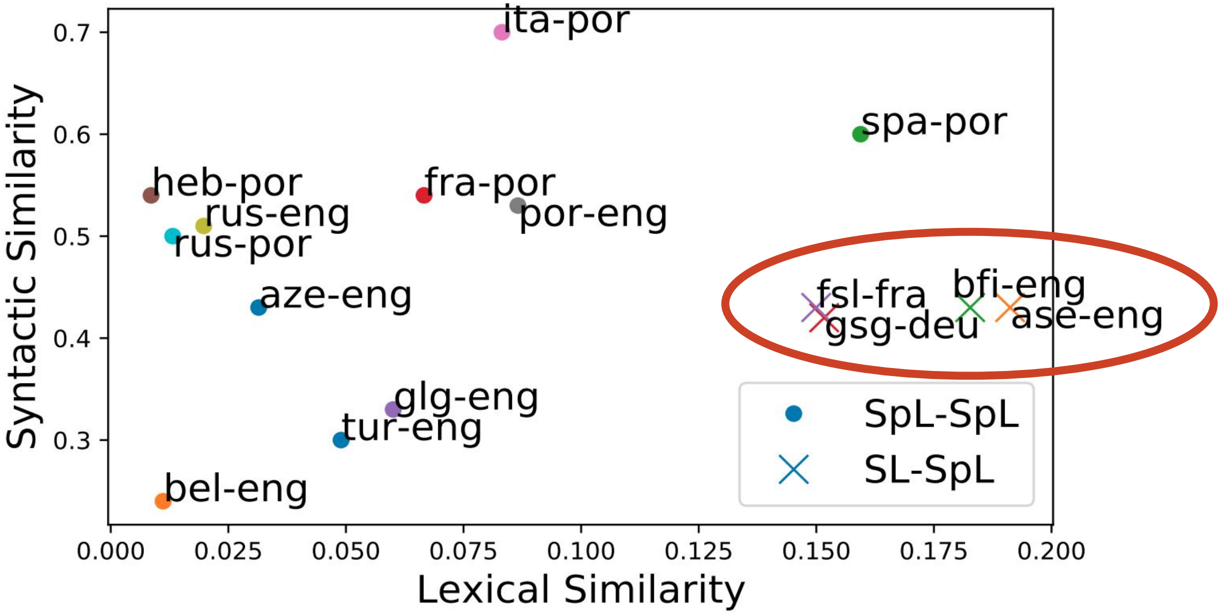
- Syntactic similarity

$$1 - d_{syn}$$

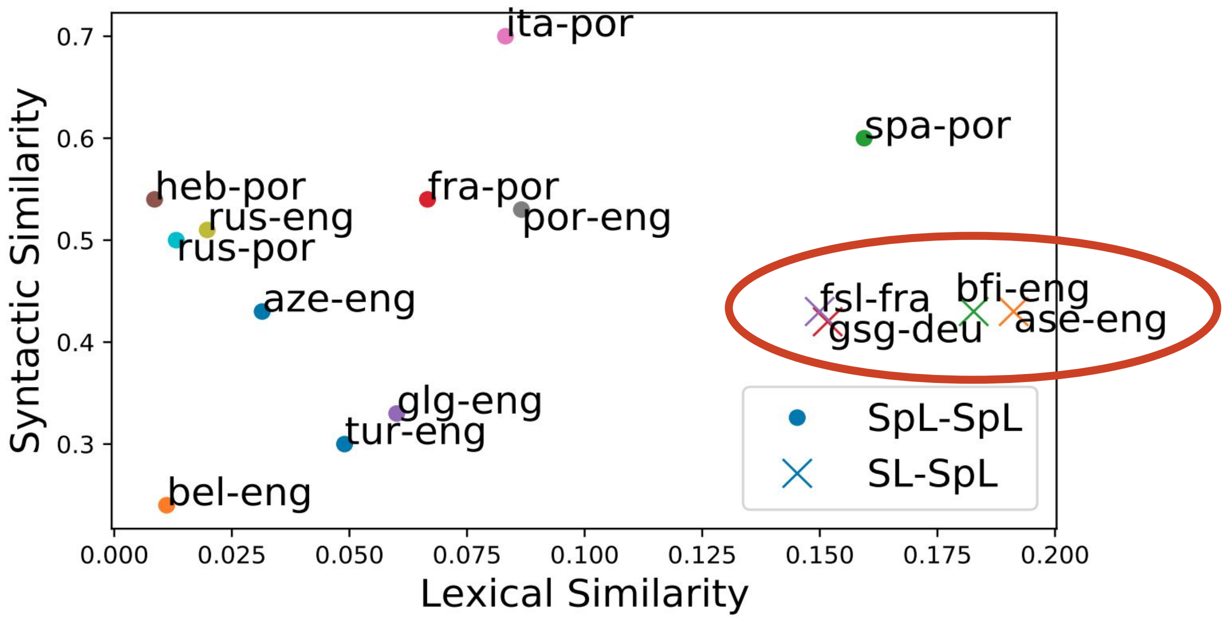
Signed vs. Spoken Languages



Signed vs. Spoken Languages



Signed vs. Spoken Languages



➔ Signed-spoken language pairs are **lexically similar** but **syntactically different**

Data Augmentation

I’m looking forward to seeing the children tomorrow.

Data Augmentation

I’m looking forward to seeing the children tomorrow.

Data Augmentation

I'm looking forward to seeing the children tomorrow.

LOOK FORWARD SEE CHILD TOMORROW

Data Augmentation

I'm looking forward to seeing the children tomorrow.

FORWARD LOOK TOMORROW CHILD SEE

Data

- **NCSLGR** (SignStream, 2007)
  - American Sign Language (ASL) - English
  - 1,875 parallel sentences

- **PHOENIX 2014T** (Camgoz et al., 2018)
  - German Sign Language (DGS) - German
  - 8,257 parallel sentences

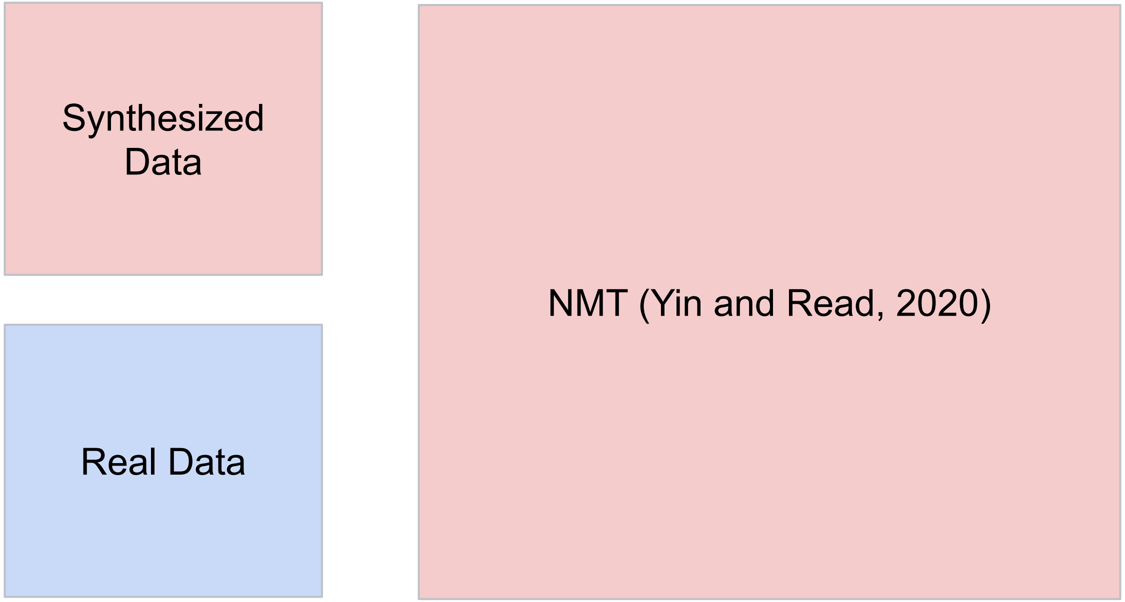
Model Training

Synthesized Data

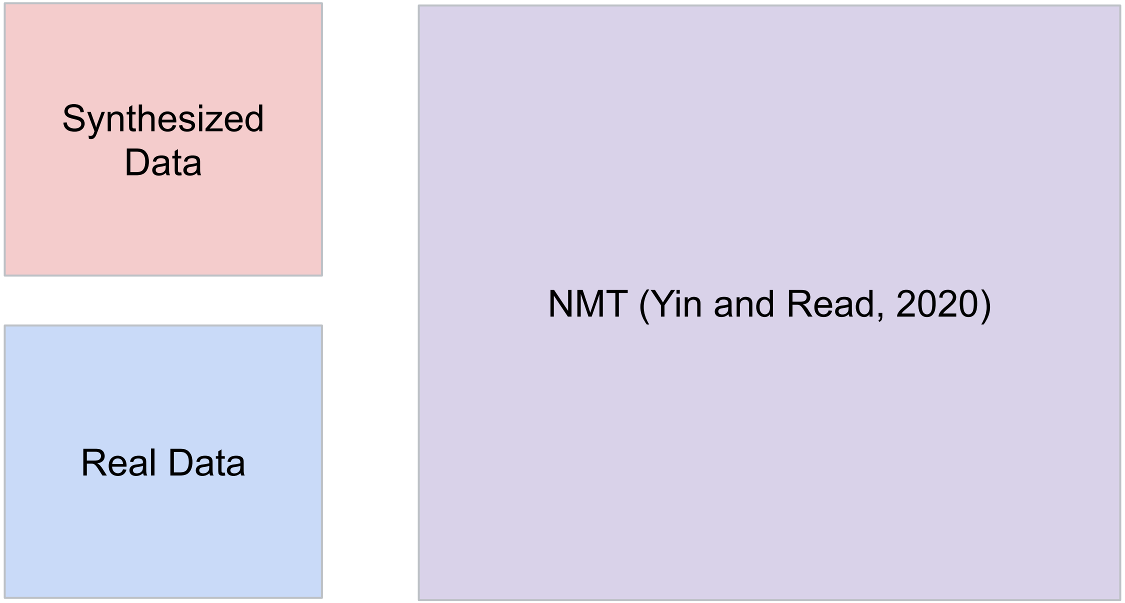
Real Data

NMT (Yin and Read, 2020)

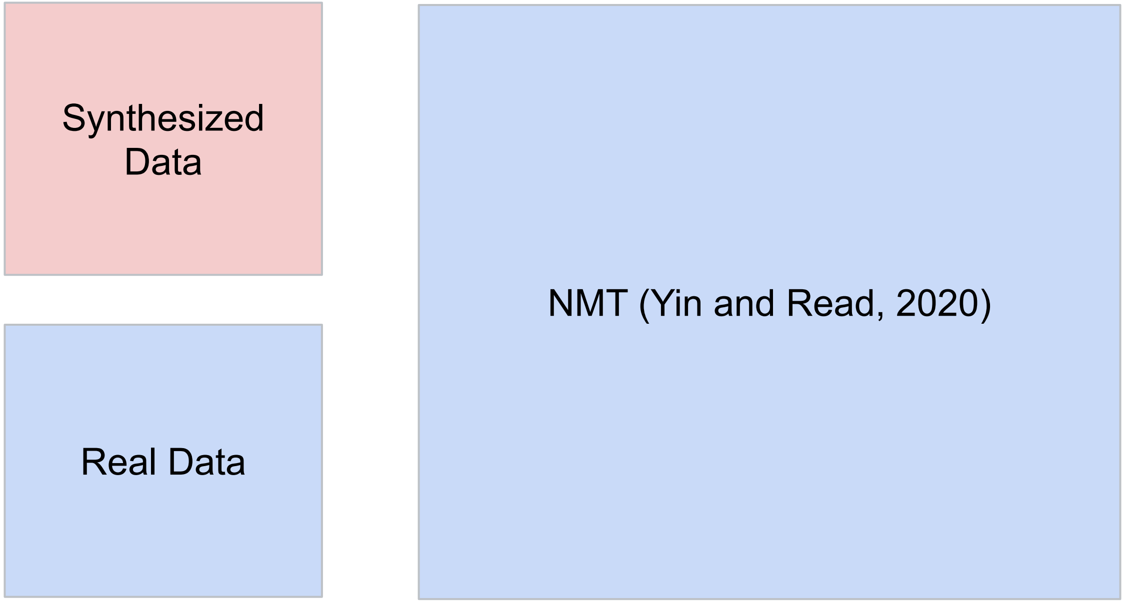
Model Training



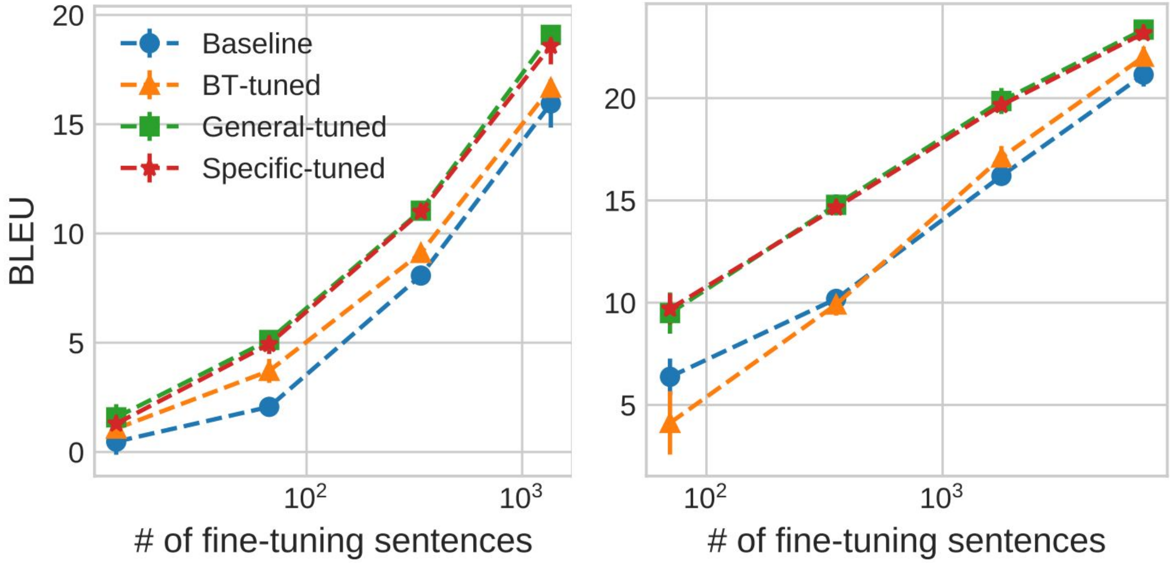
Model Training



Model Training

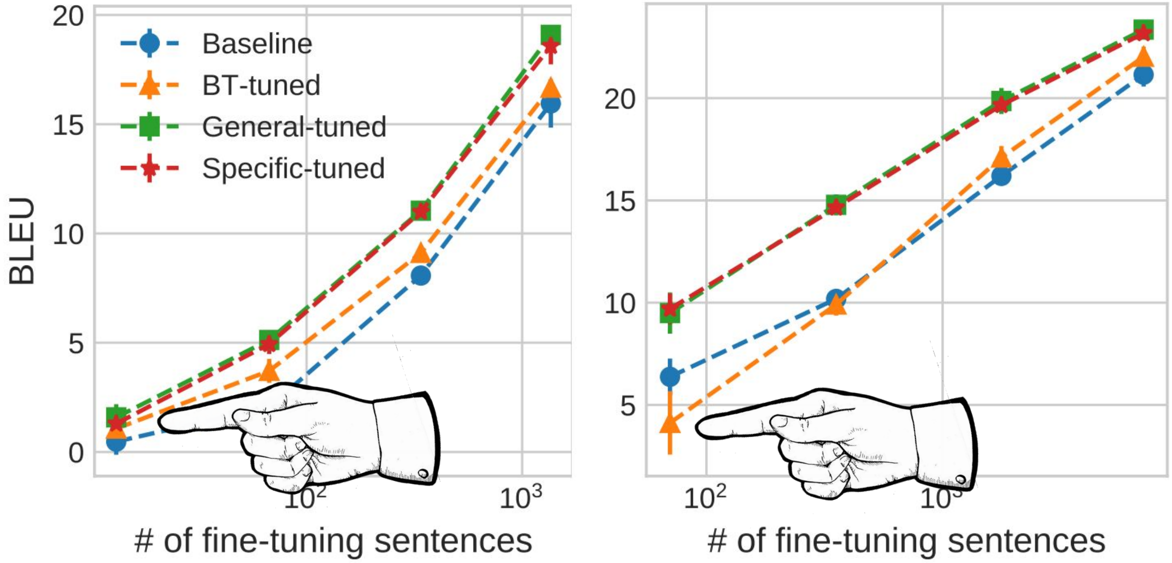


Results



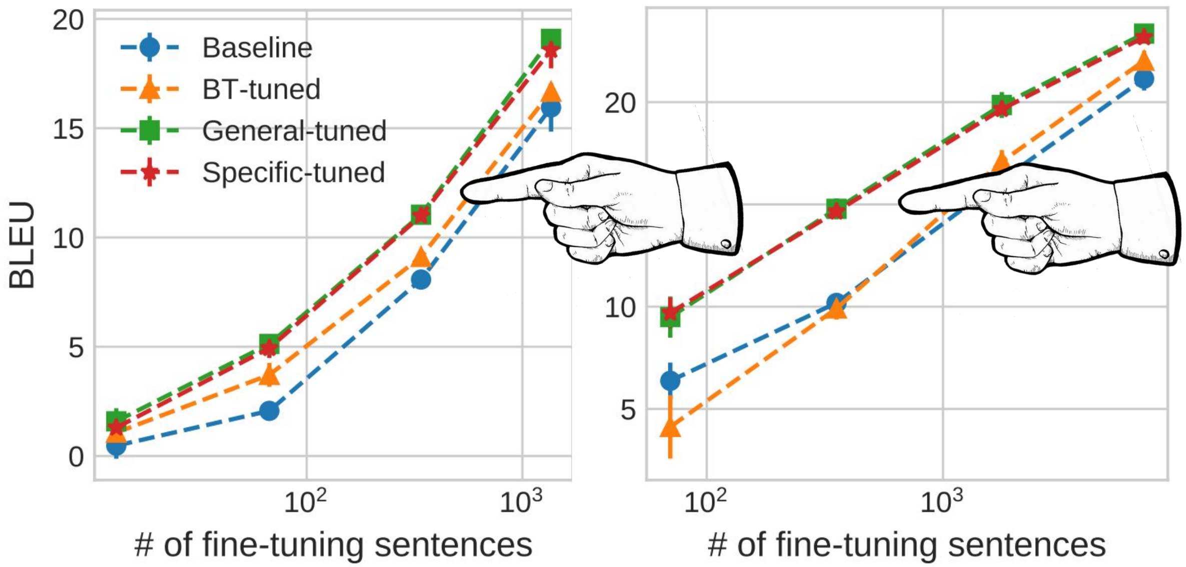
(a) NCSLGR (ASL) (b) PHOENIX (DGS)

Results



(a) NCSLGR (ASL) (b) PHOENIX (DGS)

Results



(a) NCSLGR (ASL) (b) PHOENIX (DGS)

Results

- Consistent translation improvements using **data augmentation** to leverage lexical similarities and handle syntactic differences

- Data augmentation using **monolingual spoken language data** is a promising approach

Signed Coreference Resolution

Kayo Yin, Kenneth DeHaan, Malihe Alikhani

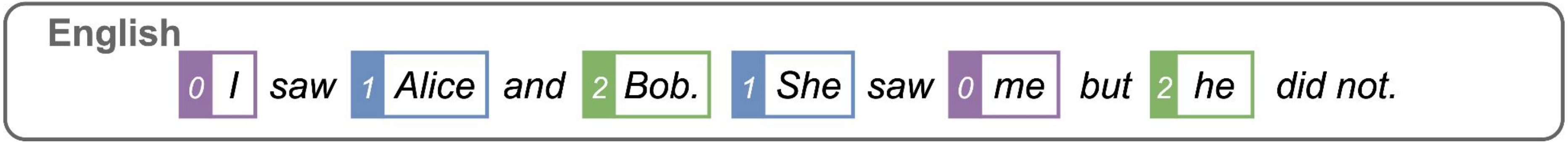
(EMNLP 2021)

Coreference Resolution

English

I saw Alice and Bob. She saw me but he did not.

Coreference Resolution



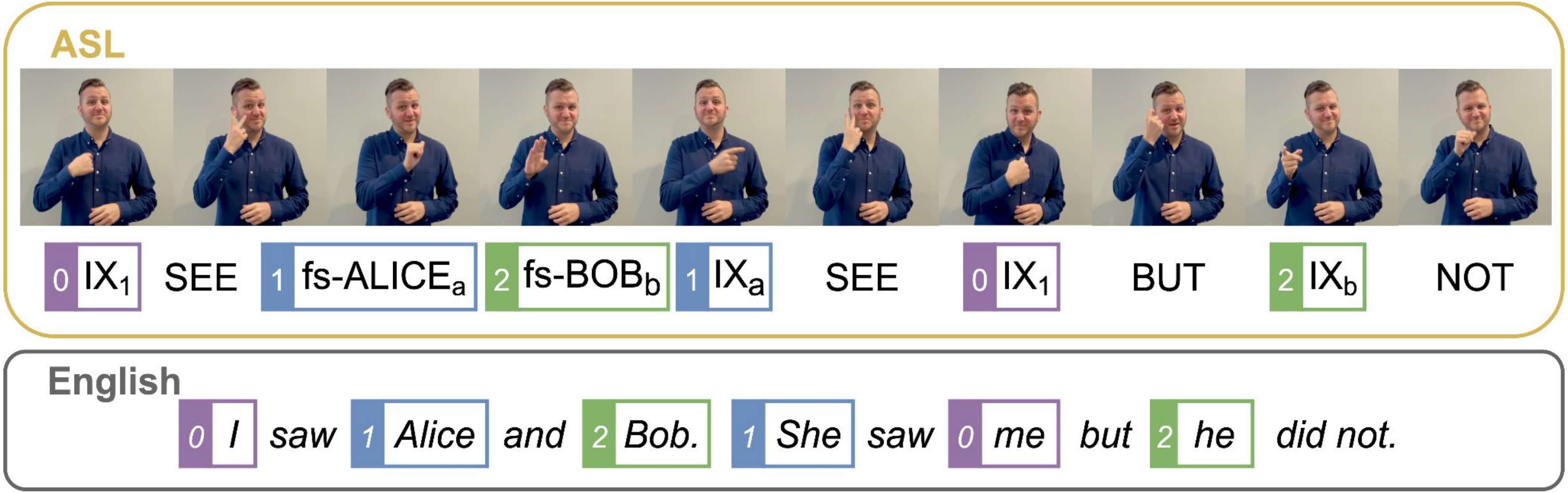
Signed Coreference Resolution

ASL

English

0 *I* *saw* 1 *Alice* *and* 2 *Bob.* 1 *She* *saw* 0 *me* *but* 2 *he* *did not.*

Signed Coreference Resolution



Signed Coreference Resolution

➔ Novel challenges in modeling **discourse** and **spatial context**

Signed Coreference Resolution

➔ Novel challenges in modeling **discourse** and **spatial context**

➔ Better understanding of **grounding** in different forms of communication

Signed Coreference Resolution

- ➔ Novel challenges in modeling **discourse** and **spatial context**
- ➔ Better understanding of **grounding** in different forms of communication
- ➔ Broaden the scope of NLP to **multiple modalities**

Signed Coreference Resolution

- ➔ Novel challenges in modeling **discourse** and **spatial context**
- ➔ Better understanding of **grounding** in different forms of communication
- ➔ Broaden the scope of NLP to **multiple modalities**
- ➔ Enable **Sign Language Processing** technologies

Outline

1. Pronominal Pointing Signs
2. Signed Coreference Resolution
3. Unsupervised Continuous Multigraph
4. Results & Discussion

Pronominal Pointing Signs

➔ Pointing signs with a **pronominal** function

Pronominal Pointing Signs

➔ Pointing signs with a **pronominal** function

➔ Referents are established in the **signing space**

Pronominal Pointing Signs

➔ Pointing signs with a **pronominal** function

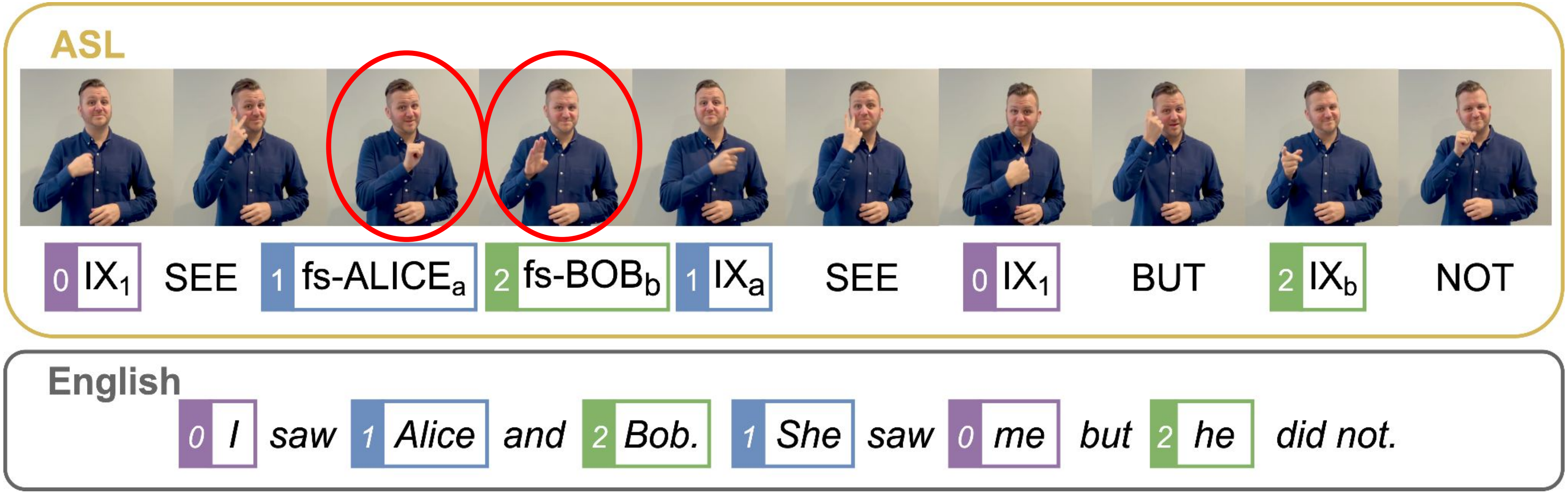
➔ Referents are established in the **signing space**

➔ Point to the **actual location** of the referent

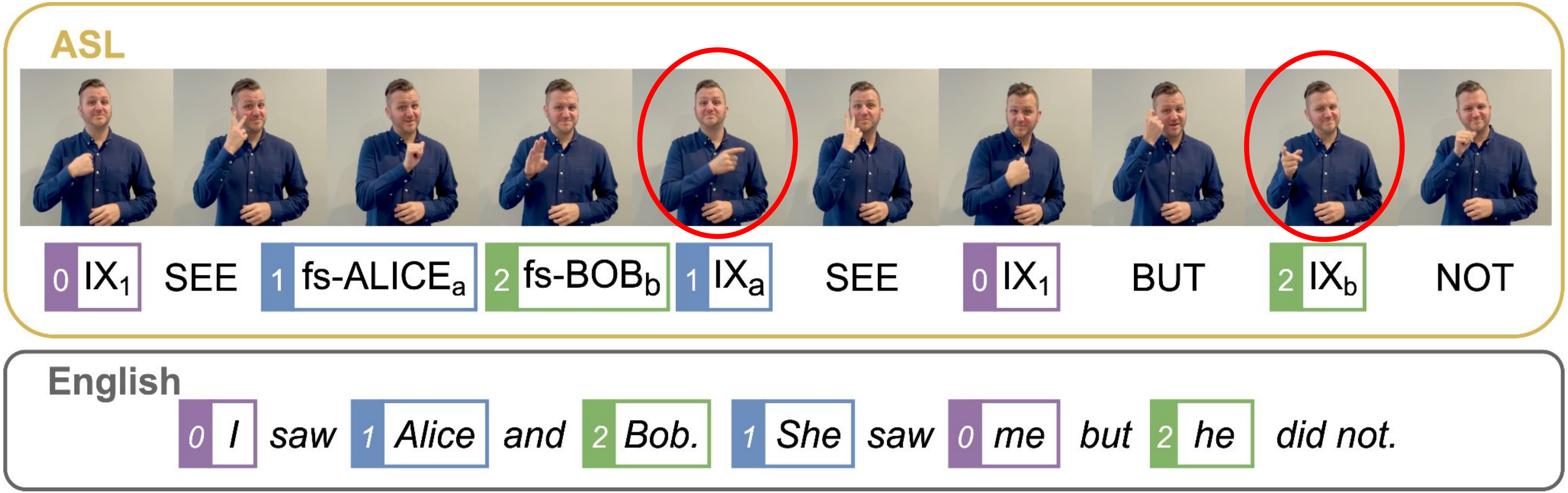
Pronominal Pointing Signs

- ➔ Pointing signs with a **pronominal** function
- ➔ Referents are established in the **signing space**
- ➔ Point to the **actual location** of the referent
- ➔ Assign a **locus** to the referent

Pronominal Pointing Signs



Pronominal Pointing Signs



Complexities of Pointing Signs

➔ Pointing signs can serve **other** functions

Complexities of Pointing Signs

➔ Pointing signs can serve **other** functions

➔ Difficult to distinguish between different pointing signs based solely on **local visual features**

Complexities of Pointing Signs

English Pronouns

ASL Pointing Signs

+ Carry some meaning on its own

Complexities of Pointing Signs

English Pronouns

+ Carry some meaning on its own

ASL Pointing Signs

- Use the same handshape, harder to distinguish on its own

Complexities of Pointing Signs

English Pronouns

+ Carry some meaning on its own
- The same word can refer to multiple entities at once

ASL Pointing Signs

- Use the same handshape, harder to distinguish on its own

Complexities of Pointing Signs

<u>English Pronouns</u>

\+ Carry some meaning on its own
\- The same word can refer to multiple entities at once

<u>ASL Pointing Signs</u>

\- Use the same handshape, harder to distinguish on its own

My mother never liked Alice, <u>she</u> thought <u>she</u> was up to no good.

Complexities of Pointing Signs

English Pronouns

+ Carry some meaning on its own
- The same word can refer to multiple entities at once

ASL Pointing Signs

- Use the same handshape, harder to distinguish on its own
+ 1 locus = 1 referent

Complexities of Pointing Signs

<u>English Pronouns</u>

+ Carry some meaning on its own
- The same word can refer to multiple entities at once

<u>ASL Pointing Signs</u>

- Use the same handshape, harder to distinguish on its own
+ 1 locus = 1 referent
- Loci can be reassigned to different referents

Complexities of Pointing Signs

<u>English Pronouns</u>

+ Carry some meaning on its own
- The same word can refer to multiple entities at once

<u>ASL Pointing Signs</u>

- Use the same handshape, harder to distinguish on its own
+ 1 locus = 1 referent
- Loci can be reassigned to different referents
- Referents can be assigned multiple loci

Why study Signed Coreference Resolution in NLP?

➔ Theories of coreference in spoken languages may be **extended** to signed languages

Why study Signed Coreference Resolution in NLP?

➔ Theories of coreference in spoken languages may be **extended** to signed languages

- ◆ Discourse Representation Theory (Kamp et al., 2011; Steinbach and Onea 2016)

Why study Signed Coreference Resolution in NLP?

- ➔ Theories of coreference in spoken languages may be **extended** to signed languages
  - ◆ Discourse Representation Theory (Kamp et al., 2011; Steinbach and Onea 2016)
  - ◆ First mention effect (Gernsbacher and Hargreaves, 1988; Wienholz et al., 2020)

Why study Signed Coreference Resolution in NLP?

- ➔ Theories of coreference in spoken languages may be **extended** to signed languages
  - ◆ Discourse Representation Theory (Kamp et al., 2011; Steinbach and Onea 2016)
  - ◆ First mention effect (Gernsbacher and Hargreaves, 1988; Wienholz et al., 2020)
- ➔ It can help us better understand **multimodal** communication

Why study Signed Coreference Resolution in NLP?

- ➔ Theories of coreference in spoken languages may be **extended** to signed languages
  - ◆ Discourse Representation Theory (Kamp et al., 2011; Steinbach and Onea 2016)
  - ◆ First mention effect (Gernsbacher and Hargreaves, 1988; Wienholz et al., 2020)
- ➔ It can help us better understand **multimodal** communication
  - ◆ Spatial iconicity and situated referents in signed languages

Why study Signed Coreference Resolution in NLP?

- ➔ Theories of coreference in spoken languages may be **extended** to signed languages
  - ◆ Discourse Representation Theory (Kamp et al., 2011; Steinbach and Onea 2016)
  - ◆ First mention effect (Gernsbacher and Hargreaves, 1988; Wienholz et al., 2020)
- ➔ It can help us better understand **multimodal** communication
  - ◆ Spatial iconicity and situated referents in signed languages
- ➔ Widen the **accessibility** of language technologies

Outline

1. Pronominal Pointing Signs
2. Signed Coreference Resolution
3. Unsupervised Continuous Multigraph
4. Results & Discussion

Signed Coreference Resolution

Signed Coreference Resolution

IX_1 SEE $fs\text{-}ALICE_a$ $fs\text{-}BOB_b$ IX_a SEE IX_1 BUT IX_b NOT

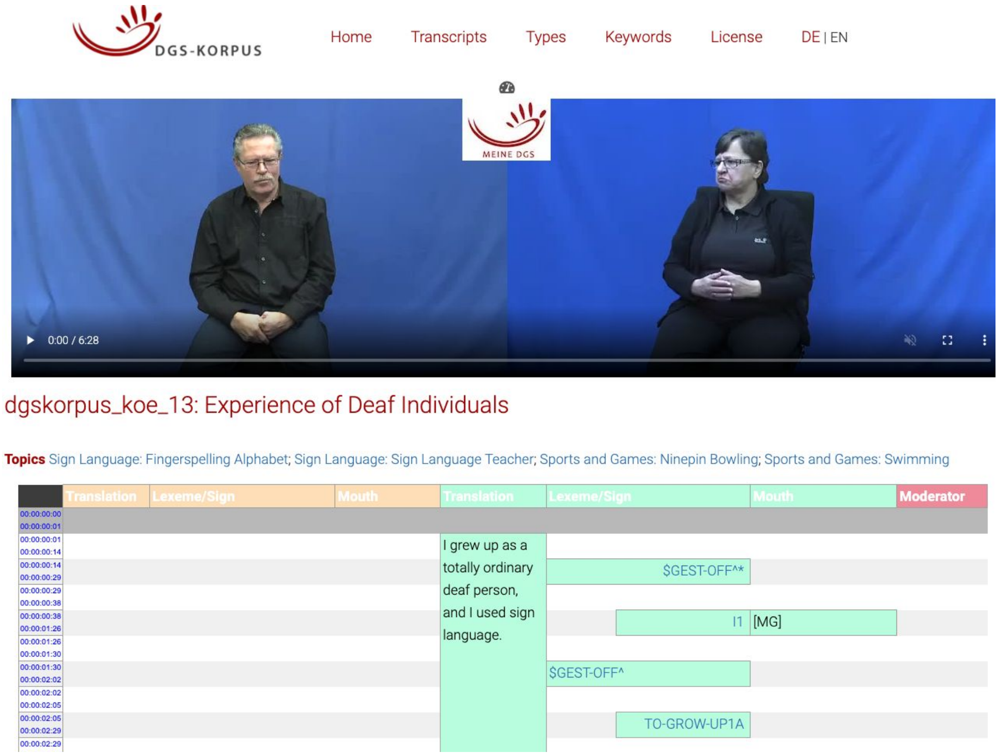
1. Mention Detection

Signed Coreference Resolution

0 IX_1 SEE 1 fs-ALICE_a 2 fs-BOB_b 1 IX_a SEE 0 IX_1 BUT 2 IX_b NOT

2. Coreference Resolution

DGS-Coref Dataset



Public DGS Corpus (Hanke et al., 2020)

DGS-Coref Dataset

Task 1 (Video b'1429737', 84) - Example 61

Video: https://www.sign-lang.uni-hamburg.de/meinedgs/html/1429737_en.html#t00053952

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

DGS-Coref Dataset

- ➔ 16m30s of signing
- ➔ 3 conversations
- ➔ 5 different signers
- ➔ 288 signed sentences
- ➔ 1,457 glosses
  - ◆ 95 <I> signs
  - ◆ 8 <YOU> signs
  - ◆ 93 <INDEX> signs

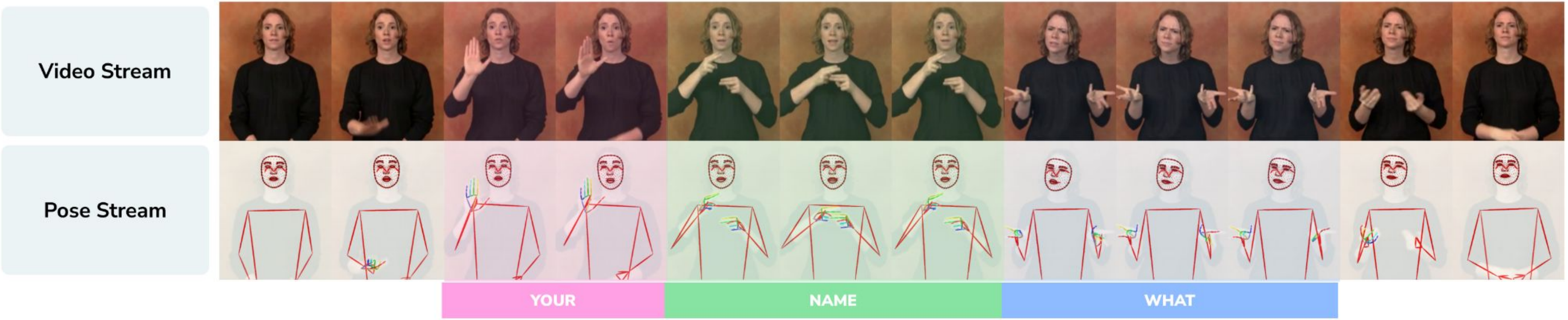
A: WITH TRIP **INDEX** SHIP **INDEX**

A: We went there with an excursion boat.

Outline

1. Pronominal Pointing Signs
2. Signed Coreference Resolution
3. Unsupervised Continuous Multigraph
4. Results & Discussion

Unsupervised Continuous Multigraph

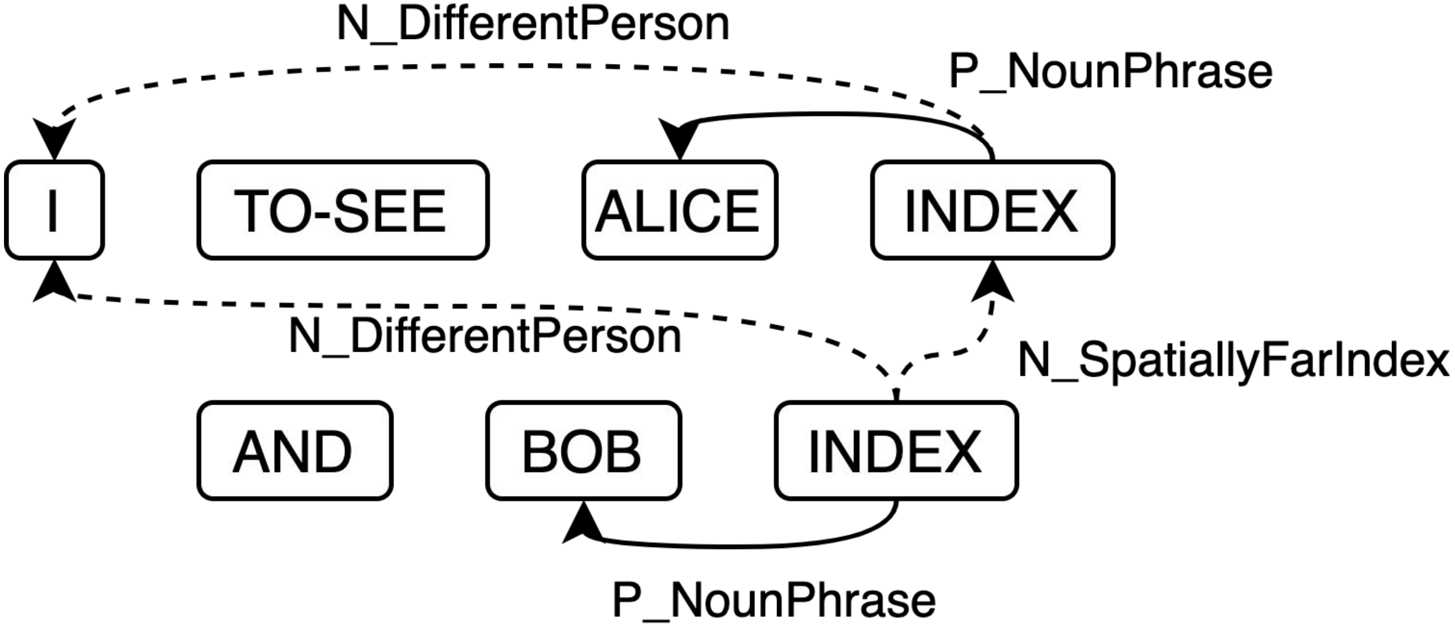


Unsupervised Continuous Multigraph

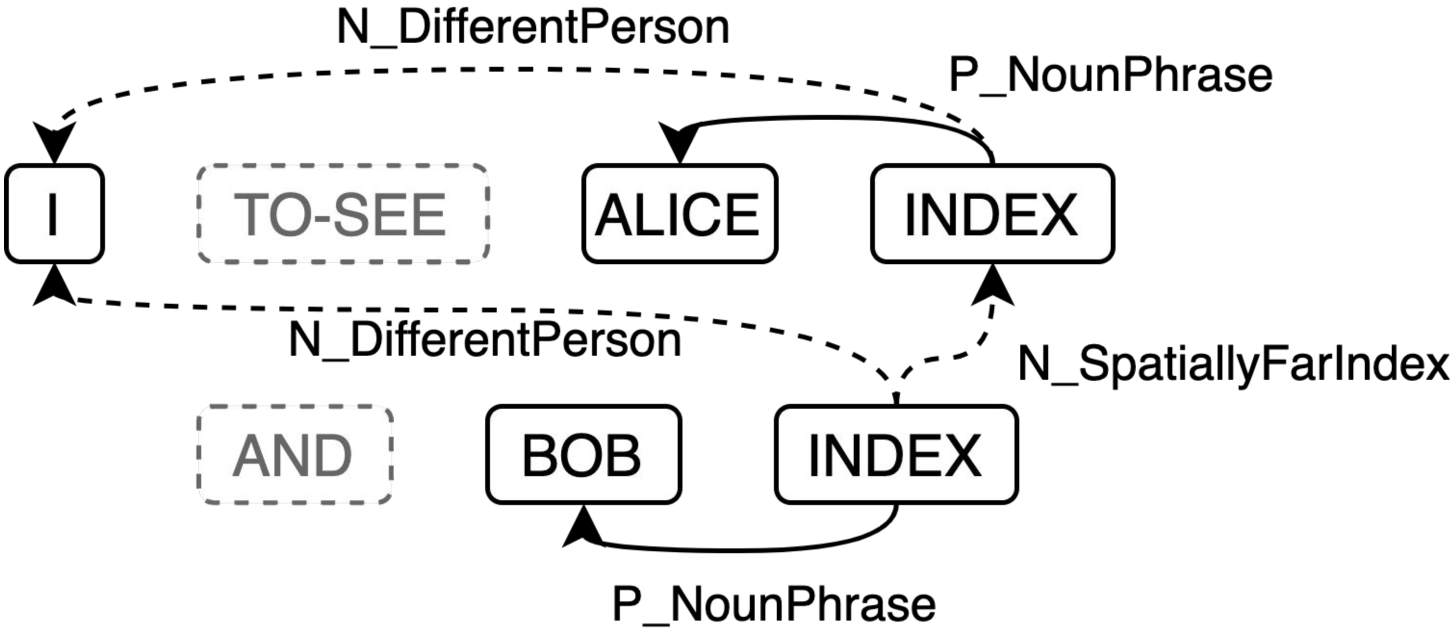
I TO-SEE ALICE INDEX

AND BOB INDEX

Unsupervised Continuous Multigraph



Unsupervised Continuous Multigraph

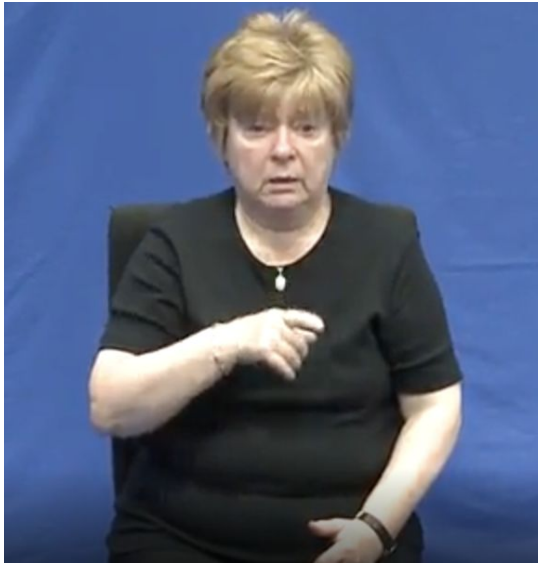


Positive Relations

1. I and I

Positive Relations

1. I and I
2. You and You

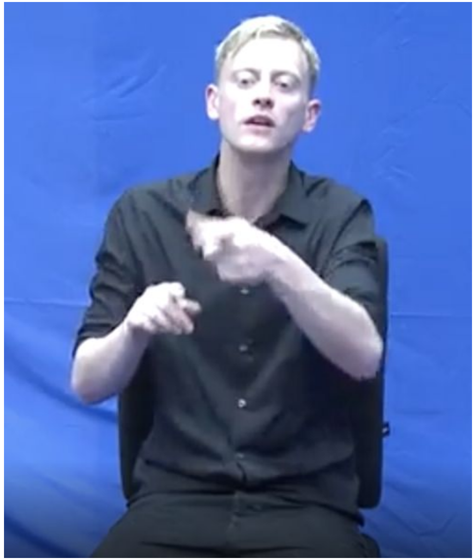


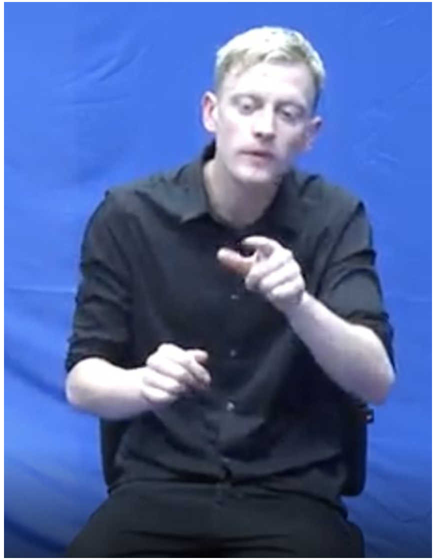
Positive Relations

1. I and I
2. You and You
3. I and You

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index



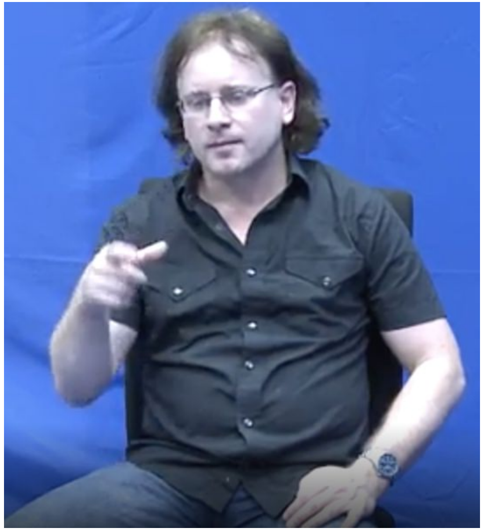


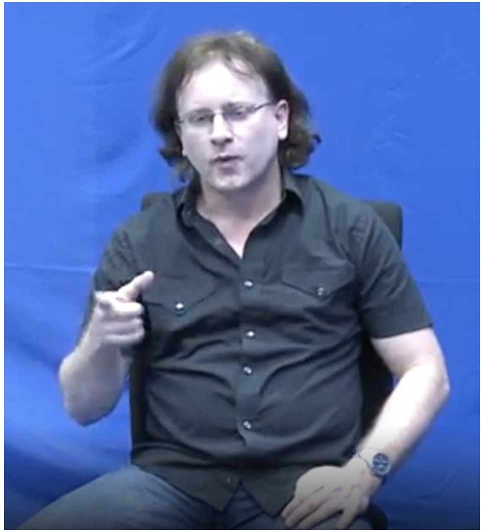
Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index



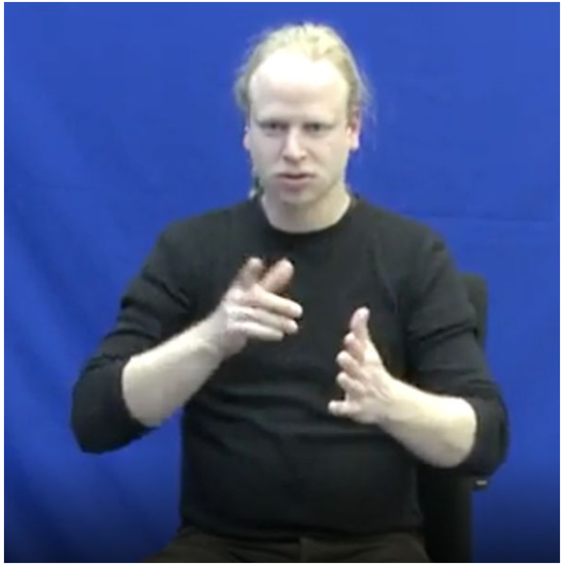


Negative Relations

1. I and I

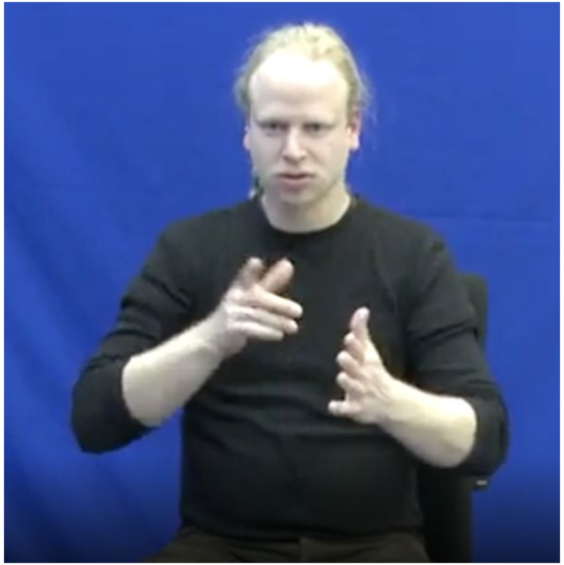
Negative Relations

1. I and I
2. You and You



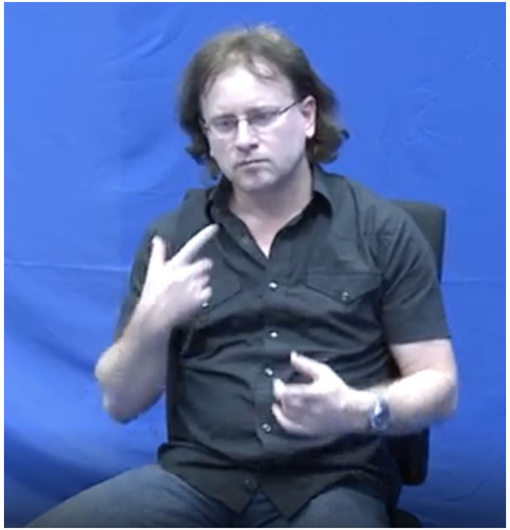
Negative Relations

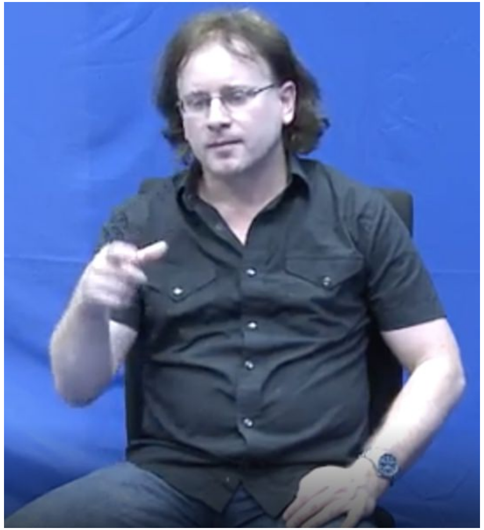
1. I and I
2. You and You
3. I and You



Negative Relations

1. I and I
2. You and You
3. I and You
4. Different Person





Negative Relations

1. I and I
2. You and You
3. I and You
4. Spatially Far Index

Weight Assignment

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index

Negative Relations

1. I and I
2. You and You
3. I and You
4. Spatially Far Index

Weight Assignment

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index

Negative Relations

$$\left.\begin{array}{l}\text{1. I and I}\\ \text{2. You and You}\\ \text{3. I and You}\\ \text{4. Spatially Far Index}\end{array}\right\}-\infty$$

Weight Assignment

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index

+0.5

Negative Relations

1. I and I
2. You and You
3. I and You
4. Spatially Far Index

$-\infty$

Weight Assignment

Positive Relations

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index

$\left.\begin{array}{l}\text{1. I and I}\\ \text{2. You and You}\\ \text{3. I and You}\end{array}\right\}$ **+0.5**

$\left.\begin{array}{l}\text{4. Temporally Close Index}\\ \text{5. Noun Phrase}\end{array}\right\}$ **+(10-t)/20**

Negative Relations

1. I and I
2. You and You
3. I and You
4. Spatially Far Index

$\left.\begin{array}{l}\text{1–4}\end{array}\right\}$ $-\infty$

Weight Assignment

<u>Positive Relations</u>

1. I and I
2. You and You
3. I and You
4. Temporally Close Index
5. Noun Phrase
6. Spatially Close Index

Items 1–3: **+0.5**

Items 4–5: **+(10-t)/20**

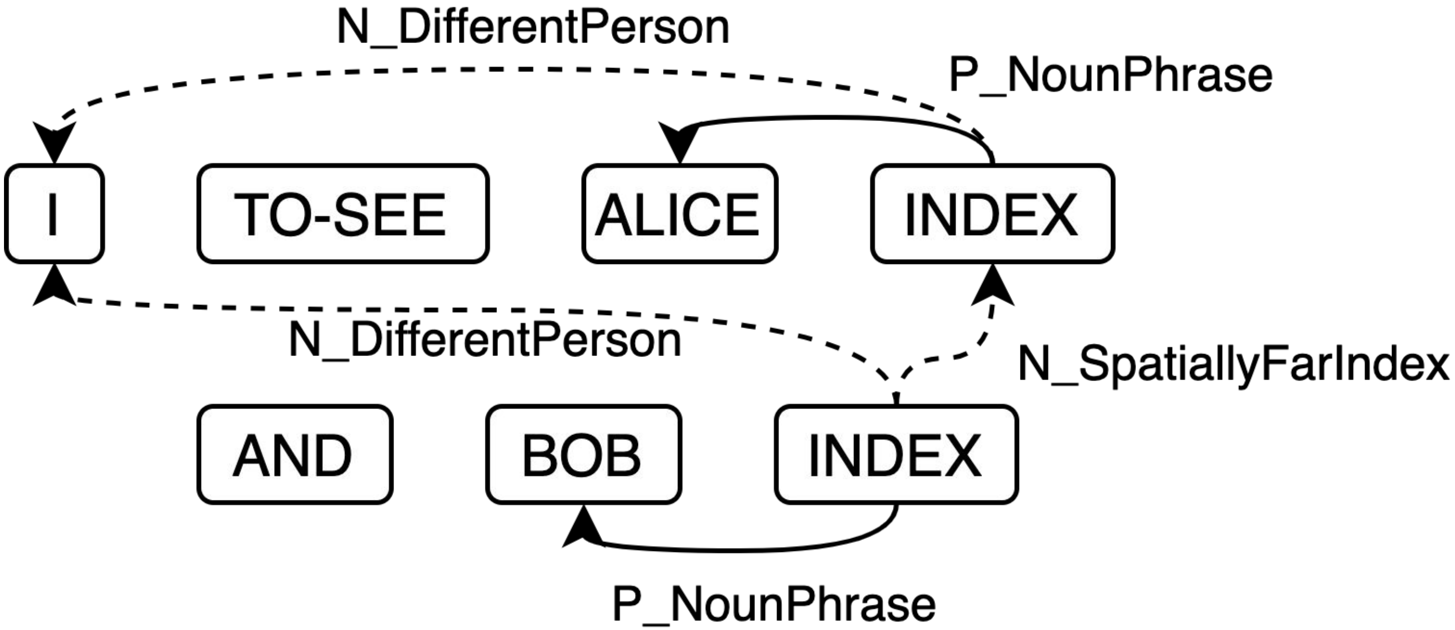
Item 6: **+(50-s)/50**

<u>Negative Relations</u>

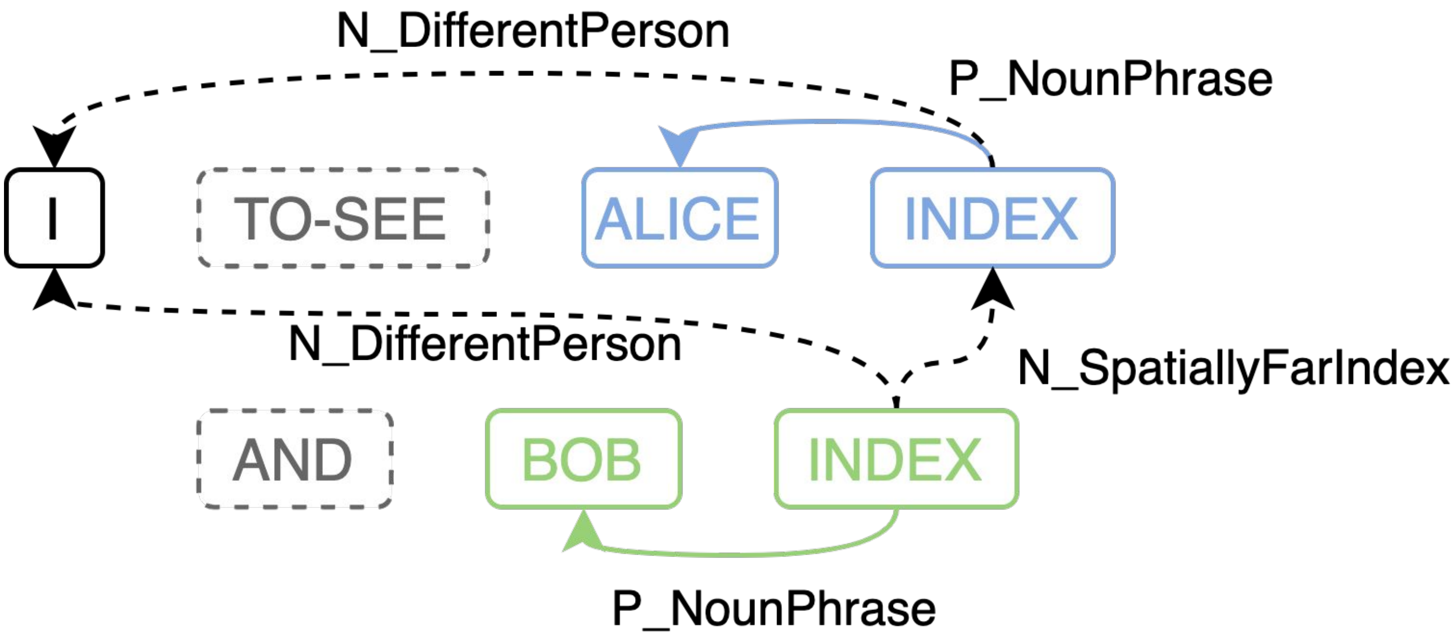
1. I and I
2. You and You
3. I and You
4. Spatially Far Index

Items 1–4: $-\infty$

Clustering


N_DifferentPerson
P_NounPhrase
I
TO-SEE
ALICE
INDEX
N_DifferentPerson
N_SpatiallyFarIndex
AND
BOB
INDEX
P_NounPhrase

Clustering

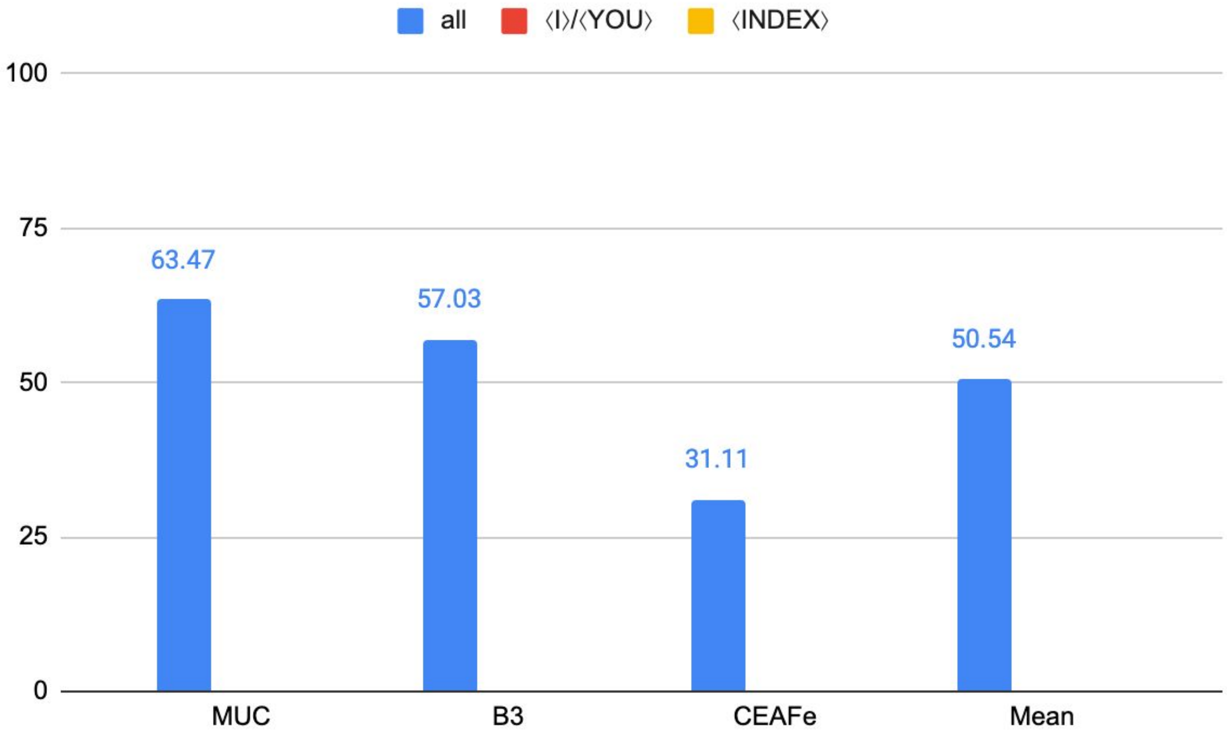


N_DifferentPerson
P_NounPhrase
I
TO-SEE
ALICE
INDEX
N_DifferentPerson
N_SpatiallyFarIndex
AND
BOB
INDEX
P_NounPhrase

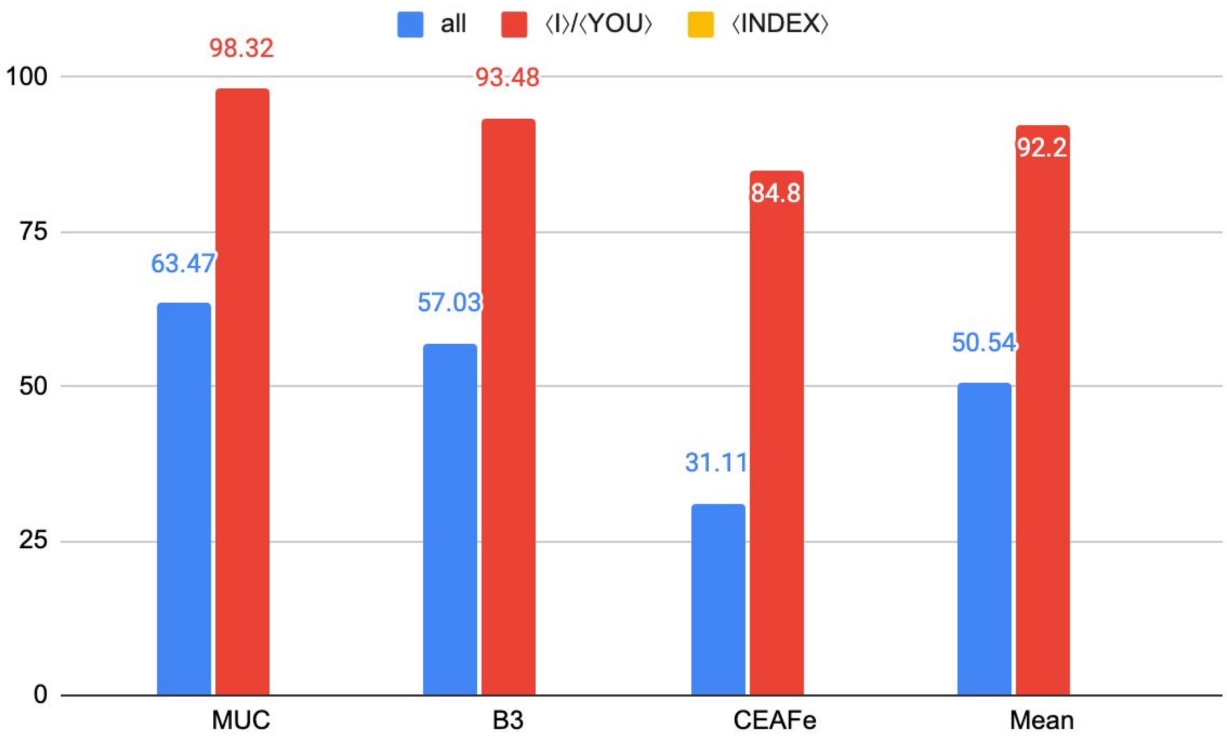
Outline

1. Pronominal Pointing Signs
2. Signed Coreference Resolution
3. Unsupervised Continuous Multigraph
4. Results & Discussion

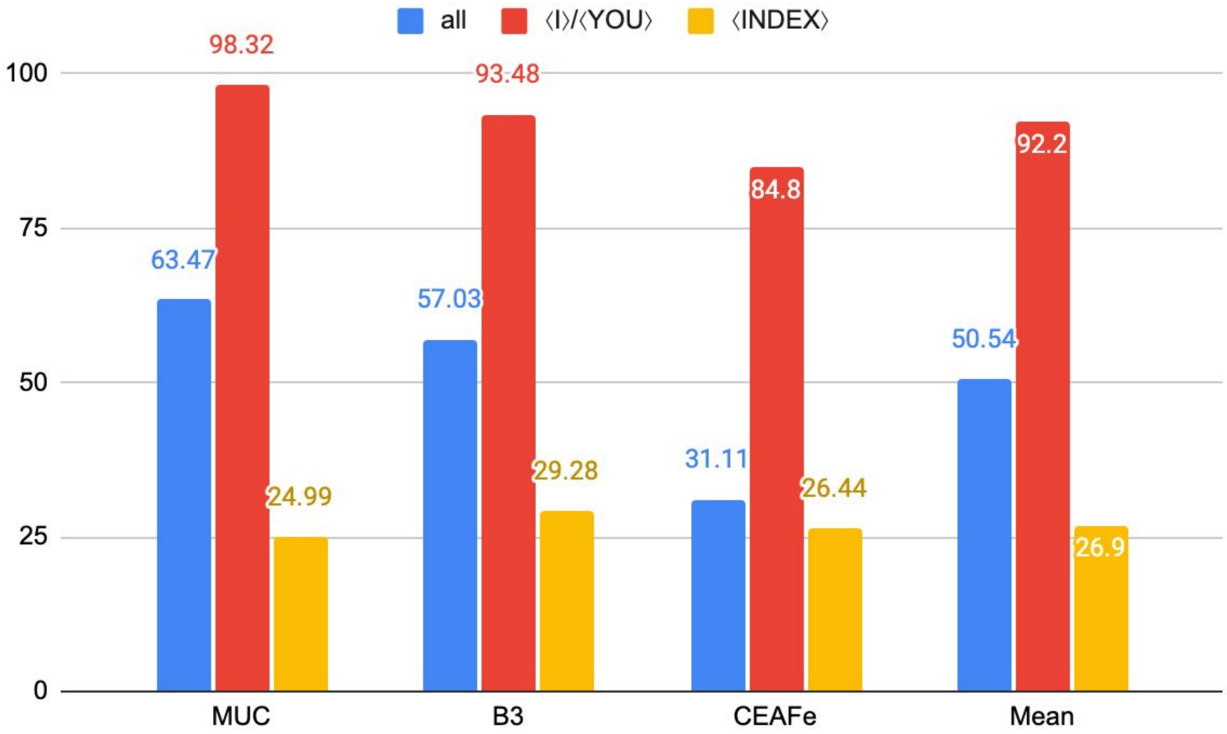
Results


all
⟨I⟩/⟨YOU⟩
⟨INDEX⟩
100
75
50
25
0
63.47
57.03
31.11
50.54
MUC
B3
CEAFe
Mean

Results

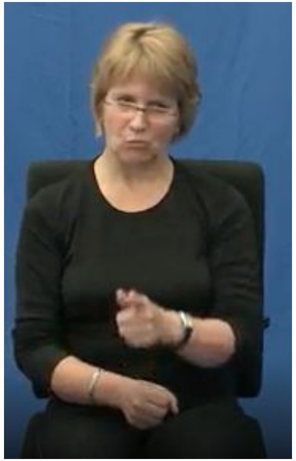

all
⟨I⟩/⟨YOU⟩
⟨INDEX⟩
100
75
50
25
0
98.32
93.48
92.2
84.8
63.47
57.03
50.54
31.11
MUC
B3
CEAFe
Mean

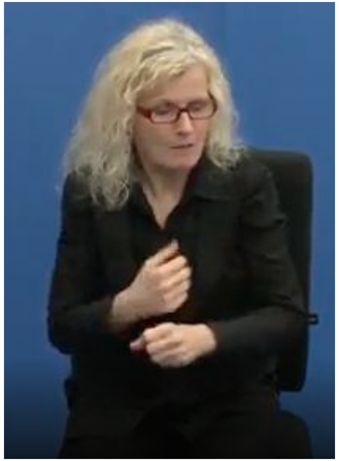
Results



all
⟨I⟩/⟨YOU⟩
⟨INDEX⟩
100
75
50
25
0
63.47
98.32
24.99
57.03
93.48
29.28
31.11
84.8
26.44
50.54
92.2
26.9
MUC
B3
CEAFe
Mean

Examples



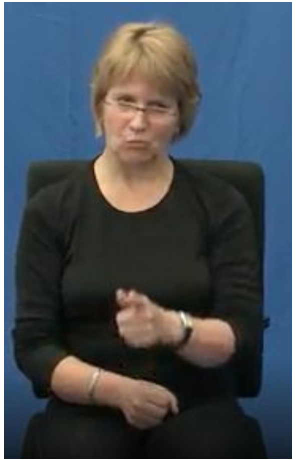
TO-SEE **YOU** GOOD **YOU**

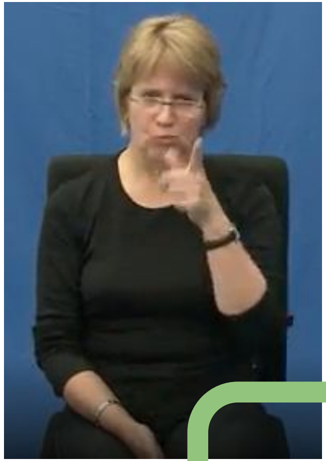
I think you could do a good job there.

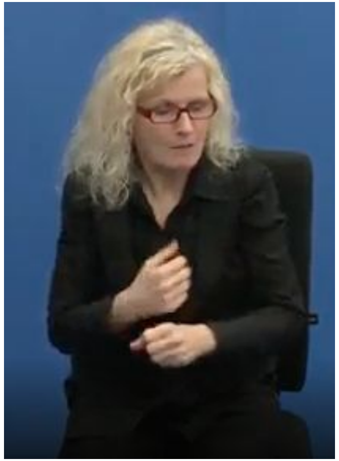
GEST-DECLINE **I** CAN NOT TO-SAY TO-HOLD-ON **I**

I can’t keep that promise

Examples







P_IAndYou

TO-SEE **YOU** GOOD **YOU**

I think you could do a good job here.

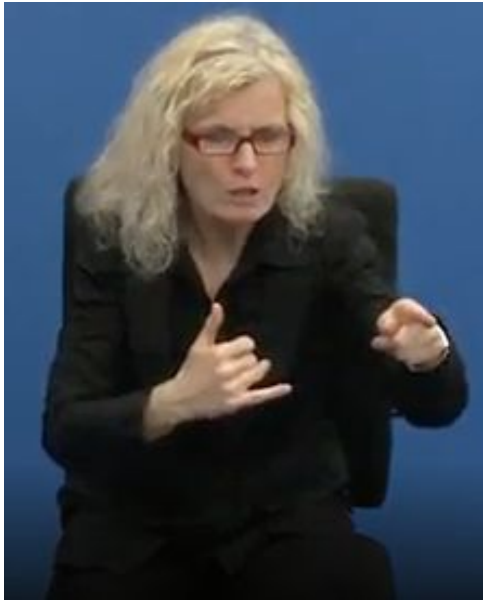
P_YouAndYou

GEST-DECLINE **I** CAN NOT TO-SAY TO-HOLD-ON **I**

I can't keep that promise

P_IAndI

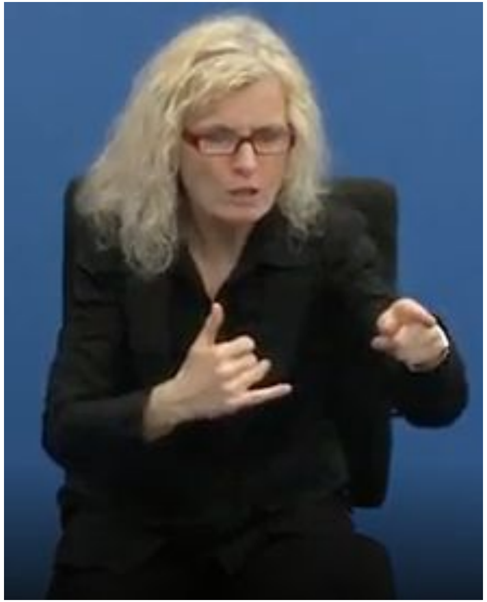
Examples



STUTTGART NUM-1 **NAME INDEX** NUM-1 FREIBURG

Once we were in Stuttgart, once in Ingolstadt and once in Freiburg.

Examples



P_NounPhrase

STUTTGART NUM-1 **NAME INDEX** NUM-1 FREIBURG

Once we were in Stuttgart, once in Ingolstadt and once in Freiburg.

Examples

WITH TRIP **INDEX** SHIP **INDEX**

We went there with an excursion boat.

Examples

P_TemporallyCloseIndex
P_SpatiallyCloseIndex

WITH TRIP **INDEX** SHIP **INDEX**

We went there with an excursion boat.

Examples

I TO-LEARN INDEX **HAMBURG INDEX**

I learned it in Hamburg.

Examples

P_TemporallyCloseIndex
P_SpatiallyCloseIndex

I TO-LEARN **INDEX** **HAMBURG** **INDEX**

I learned it in Hamburg.

Conclusion

- Signed language data is **scarce** and **hard to obtain**

Conclusion

- Signed language data is **scarce** and **hard to obtain**
  - **Data augmentation** from **monolingual spoken language data** is one promising way to mitigate this

Conclusion

- Signed language data is **scarce** and **hard to obtain**
  - **Data augmentation** from **monolingual spoken language data** is one promising way to mitigate this

- The meaning of certain signs rely on **spatial context**

Conclusion

- Signed language data is **scarce** and **hard to obtain**
  - **Data augmentation** from **monolingual spoken language data** is one promising way to mitigate this

- The meaning of certain signs rely on **spatial context**
  - **Signed Coreference Resolution** as a new challenge
  - **Unsupervised Continuous Multigraph** for SCR

Future Work

- **Pre-training** the target side decoder with spoken language data

Future Work

- **Pre-training** the target side decoder with spoken language data

- Resolve other types of **ambiguous** signs

Summary

➔ New challenge: **Signed Coreference Resolution**

➔ **Annotation** software & **DGS-Coref dataset**

Summary

➔ New challenge: **Signed Coreference Resolution**

➔ **Annotation** software & **DGS-Coref dataset**

➔ **Unsupervised Continuous Multigraph** for SCR

Summary

- ➔ New challenge: **Signed Coreference Resolution**
- ➔ **Annotation** software & **DGS-Coref dataset**
- ➔ **Unsupervised Continuous Multigraph** for SCR
- ➔ Code & data: **`https://github.com/kayoyin/scr`**

Future Work

- ➔ Detect **reassignment** of loci
- ➔ Detect **different functions** of indexing signs
- ➔ Keep track of the **dynamic** signing space
- ➔ Directly process **videos**
- ➔ Resolve other types of **pronominal** signs
- ➔ Resolve other types of **ambiguous** signs