



# Sign Language MT

Sara Morrissey



# Overview

- Introduction
- Irish Sign Language
- Problems for SLMT
- SLMT Data
- MaTrEx for SLs
- Future work



# Introduction (1)

## Motivation

- SLs are poorly resourced and lack political, social and educational recognition
- Deaf community forced to communicate in a language not natural to them
- Literacy rates of Deaf adults moderate to low, similar to 10-year-old
- MT could help alleviate communication problems



# Introduction (2)

## Research Question

Is it possible to exploit example-based machine translation techniques to facilitate the development of a complete spoken language to sign language machine translation system?



# Irish Sign Language

- ~150 years old
- Related to French Sign Language
- Developed own grammatical structure and vocabulary and reflect culture
- Dominant and preferred language of ~5,000 members of deaf community
- Irish Deaf Association, National Association for the Deaf, Centre for Deaf Studies



# Problems for SLMT

- Data
  - Lack of appropriate data
    - Subject matter
    - Format
  - Sparse data
    - Max. ~600 sentences
- Evaluation
  - Lack of gold standard
  - Manual assessment



# SLMT Data (1)

## Data Resource

- Flight information domain
  - 595 sentences: Air Travel Information System (ATIS) corpus

		EN	DE	ISL	DGS
Train	no. sentences	418			
	no. running words	3008	3544	3028	2980
	vocab. size	292	327	265	244
	no. singletons	97	118	71	84
Dev	no. sentences	59			
	no. running words	429	503	431	434
	vocab. size	134	142	131	119
Train	no. sentences	118			
	no. running words	999	856	874	877
	vocab. size	174	158	148	135
	trigram perplexity	15.7	12.4	28.3	11.39
	out of vocab.	22	22	30	15



# SLMT Data (2)

- ISL native signers signed sentences
- ELAN toolkit used to manually annotate data

The screenshot shows the ELAN software interface. The top window displays a video of a male signer in a black shirt. The interface includes a menu bar (File, Edit, Search, View, Options, Help) and a toolbar with various playback and editing controls. A timeline at the bottom shows a video segment from 00:03:02.379 to 00:03:10.829. The English translation of the video is: "I would like to find a flight that goes from Dublin to Cork, I would like the flight to be on Aer Lingus and arrive around ten p.m." Below the translation, a table of annotations is visible, with columns for time and text. The annotations are as follows:

Time	Annotation
00:03:03.000	LIKE
00:03:03.500	FIN
00:03:04.000	FLIGHT
00:03:04.500	F
00:03:05.000	DUBLIN
00:03:05.500	T
00:03:06.000	CORK
00:03:07.000	LIKE
00:03:07.500	AER LINGUS
00:03:08.000	FLIGHT (
00:03:08.500	AROU
00:03:09.000	T
00:03:09.500	CLS: GR
00:03:10.000	T





# SLMT Data (3)

## Annotation

- Transcribing information from video data
- Subjective: annotator decides level of detail
- Categories can include gloss term, NMF detail, repetition or location detail
- Easy alignment with timeline and other language tiers – suitable for recognition
- Example: *Is flight BA a round trip?*  
IX-FLIGHT FLIGHT BA ROUND TRIP IX-FLIGHT palm-up



# MaTrEx for SLs <sup>(1)</sup>

- DCU MaTrEx MT system (Nicolas' talk)
- ISL to English
- Translation process
  - Input: ISL annotated sentence
  - Search for matches on sentential and sub-sentential level on source side and retain corresponding English translations from target
  - Recombine resulting units to form English sentence for output



# MaTrEx for SIs (2)

- MaTrEx Experiments
- 4 language pairs:
  - ISL → English *and* ISL → German
  - DGS → English *and* DGS → German
- Baseline: basic SMT MaTrEx model
- EBMT chunks
  - Marker-based chunks (T1)
  - Combined format chunks (T2)
- Distortion limit (DL)



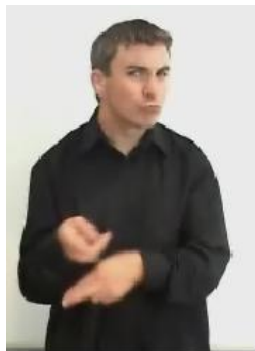
# Experiment Results

		BLEU	WER	PER
ISL-EN	baseline	51.63	39.32	29.79
	<i>DL = 10</i>	<i>52.18</i>	<i>38.48</i>	<i>29.67</i>
	T1 chunks	50.69	37.75	30.76
	<i>DL = 10</i>	<i>51.31</i>	<i>37.39</i>	<i>30.63</i>
	T2 chunks	49.76	39.92	32.44
<i>DL = 10</i>	<i>50.32</i>	<i>39.56</i>	<i>32.32</i>	
ISL-DE	baseline	38.18	48.52	38.79
	<i>DL = 10</i>	<i>39.69</i>	<i>47.25</i>	<i>38.47</i>
	T1 chunks	40.67	46.72	38.58
	<i>DL = 10</i>	<i>42.13</i>	<i>45.45</i>	<i>38.16</i>
	T2 chunks	38.54	46.93	38.05
<i>DL = 10</i>	<i>40.09</i>	<i>45.66</i>	<i>37.63</i>	
DGS-EN	baseline	45.25	48.85	32.08
	<i>DL = 10</i>	<i>48.40</i>	<i>41.37</i>	<i>30.88</i>
	T1 chunks	44.74	50.66	31.72
	<i>DL = 10</i>	<i>47.22</i>	<i>44.14</i>	<i>31.12</i>
	T2 chunks	44.34	49.93	33.17
<i>DL = 10</i>	<i>47.43</i>	<i>42.82</i>	<i>32.20</i>	
DGS-DE	baseline	38.66	55.28	39.53
	<i>DL = 10</i>	<i>42.09</i>	<i>50.31</i>	<i>39.53</i>
	T1 chunks	34.86	56.65	39.53
	<i>DL = 10</i>	<i>39.38</i>	<i>51.37</i>	<i>38.79</i>
	T2 chunks	35.63	55.81	39.74
<i>DL = 10</i>	<i>40.29</i>	<i>50.31</i>	<i>38.90</i>	

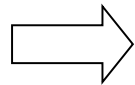


# Future Work (1)

- To date: annotation to text
- Next: SL recognition to text



Sign for 'find'



Annotation

MaTrEx Translation



*find*

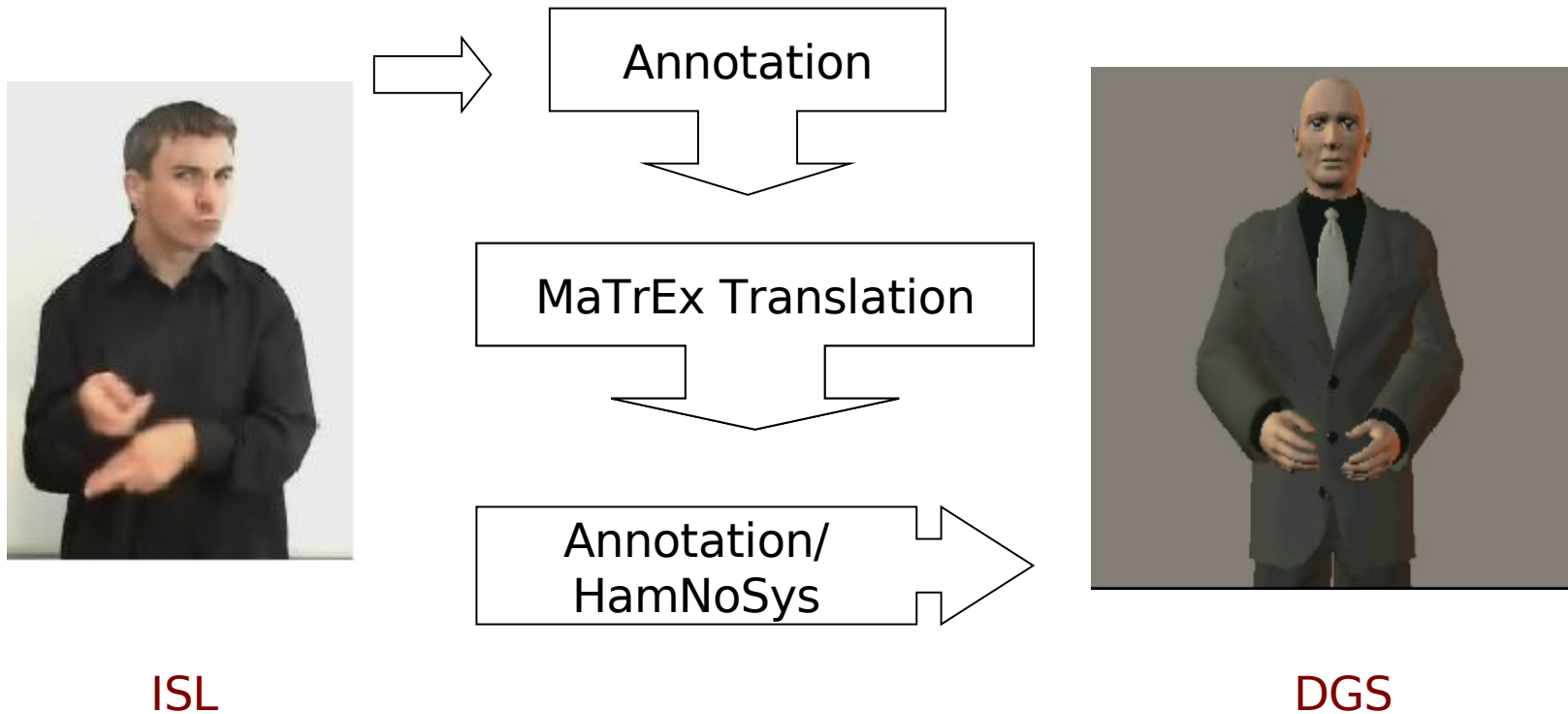
– Reverse language direction:

- text to annotation/SL with avatar output



# Future Work (2)

- SL to SL (ISL to DGS)





# Future Work <sup>(3)</sup>

- Issues of evaluation
  - Manual evaluation for SL output
  - What about internal translation stages?
- Decision of using deeper annotation level to include phonemes or HamNoSys

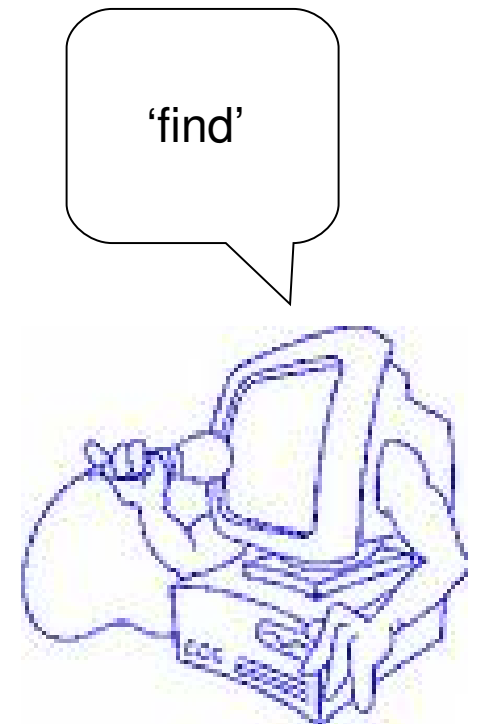
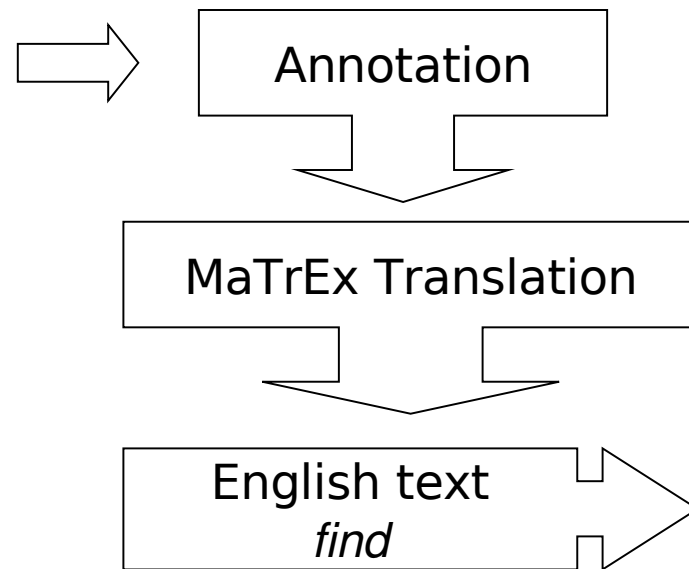


# Future Work (3)

- Speech Synthesis research group in University College Dublin (UCD)
- SL to speech system



Sign for 'find'







# Publications

- **Sara Morrissey** and Andy Way. 2007. Joining Hands: Developing a Sign Language Machine Translation System with and for the Deaf Community. In *Proceedings of the Conference and Workshop on Assistive Technology for People with Vision and Hearing Impairments*. Granada, Spain, to appear.
- **Sara Morrissey**, Andy Way, Daniel Stein, Jan Bungeroth, Hermann Ney. 2007. Towards a Hybrid Data-Driven MT System for Sign Language Translation. In *Proceedings of Workshop Example-Based Machine Translation (MT XI - 07)*, Copenhagen, Denmark, to appear
- **Sara Morrissey**. 2006. Experiments in Sign Language Machine Translation Using Examples. In *Proceedings of IBM CASCON Dublin Symposium at the 16th Annual International Conference on Computer Science and Software Engineering*. Dublin, Ireland (to appear).
- **Sara Morrissey** and Andy Way. 2006. Lost in Translation: the Problems of Using Mainstream MT Evaluation Metrics for Sign Language Translation. In *Proceedings of Strategies for developing machine translation for minority languages: 5th SALTMIL Workshop on Minority Languages*. Genoa, Italy. pp.91-98
- Armstrong, S., D. Groves, M. Flanagan, Y.Graham, B. Mellebeek, **S. Morrissey**, N. Stroppa & A. Way (2006) The MaTreX System: Machine Translation Using Examples.
- **Sara Morrissey** and Andy Way. 2005. An Example-Based Approach to Translating Sign Language. In *Proceedings of Workshop Example-Based Machine Translation (MT X - 05)*, Phuket, Thailand, pp.109-116
- **Sara Morrissey**. 2005. An Example-based Approach to the Machine Translation of Sign Languages. In *Proceedings of IBM CASCON Dublin Symposium*. Dublin, Ireland.



# Thank You



Questions or comments?