

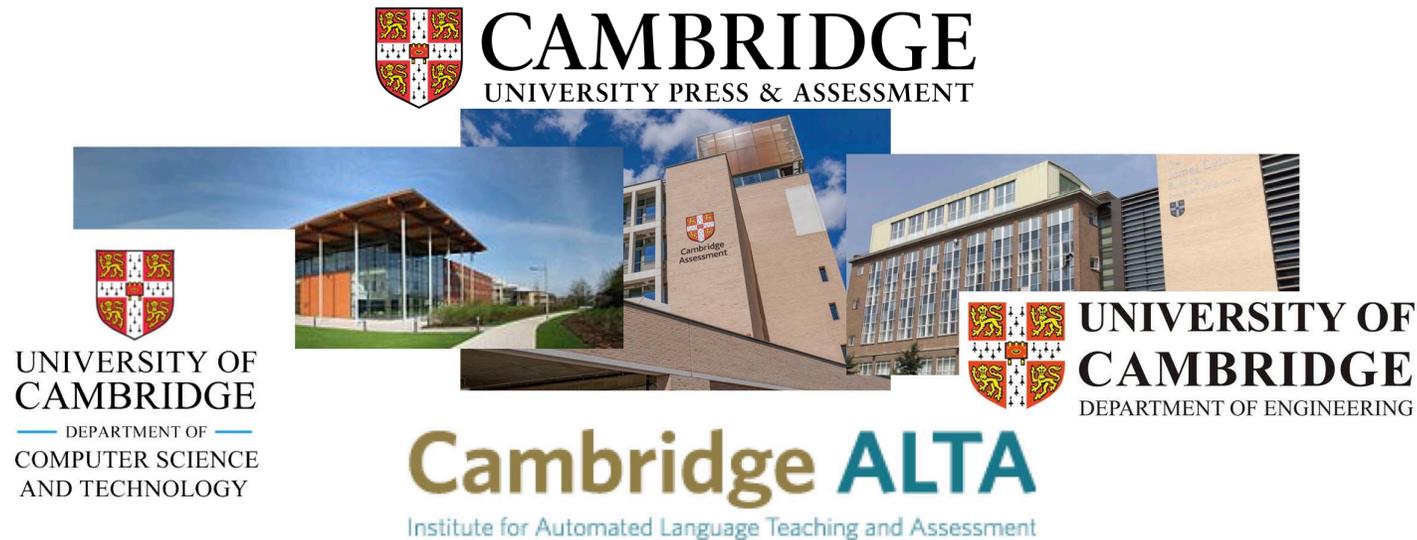
# Automated Assessment and Feedback: the Role of Spoken Grammatical Error Correction

Kate Knill

ALTA Institute, Cambridge University Engineering Department

14 December 2022

# Cambridge Automated Language Teaching and Assessment Institute



- Virtual Institute for cutting-edge research on second language (L2) English assessment
  - Machine Learning and Natural Language Processing
  - Develop technology to enhance assessment and learning
  - Look to benefit learners and teachers worldwide



Prof Mark Gales



Yiting "Edie" Lu



Stefano Bannò

Plus members of the ALTA Spoken Language Processing Technology Project,  
Cambridge University Engineering Department

# Talk Outline

- Automated Learning and Assessment
- Spoken Grammatical Error Correction (GEC)
- Systems for Spoken GEC
- Using Spoken GEC in Automatic Assessment
- Using Spoken GEC for Learner/Teacher Feedback
- Open Challenges

# Growth in Automated Learning and Assessment

Linguaskill ▶▶



VIP KID



Upskill ▶▶  
from Cambridge



Pearson



ENGLISH  
aula.com



duolingo  
english test



duolingo

Liulishuo

+Babbel

ELSA



UNIVERSITY OF  
CAMBRIDGE

Cambridge ALTA  
Institute for Automated Language Teaching and Assessment

# Automated Learning and Assessment for English L2

- Close to 2 billion people worldwide use and/or are learning English as a second language
  - Not enough teachers or examiners!
  - Automated assessment and Computer-Aided Language Learning (CALL) systems
  - Data-driven Integrated Learning and Assessment
- Speaking is key skill for communication
  - Many systems ignore or heavily restrict speech input – not testing communication
  - **Free speaking**
    - **Monologic e.g. giving a presentation**
    - Dialogic e.g. discussing at a conference

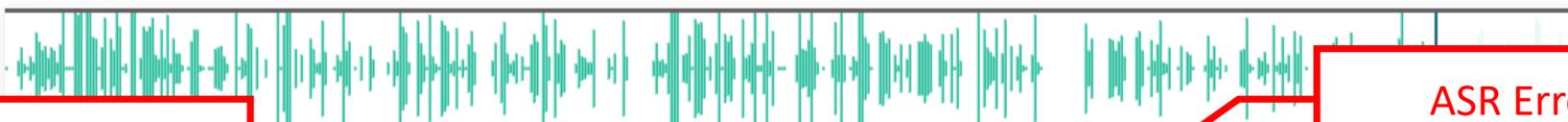
# L2 learner speech data is challenging!

Answer



Long turn 1

Talk about a training course you attended for your work. You should say: • what the course was about • why you went on the course • what you learnt from it.



No punctuation/sentences

ASR Errors

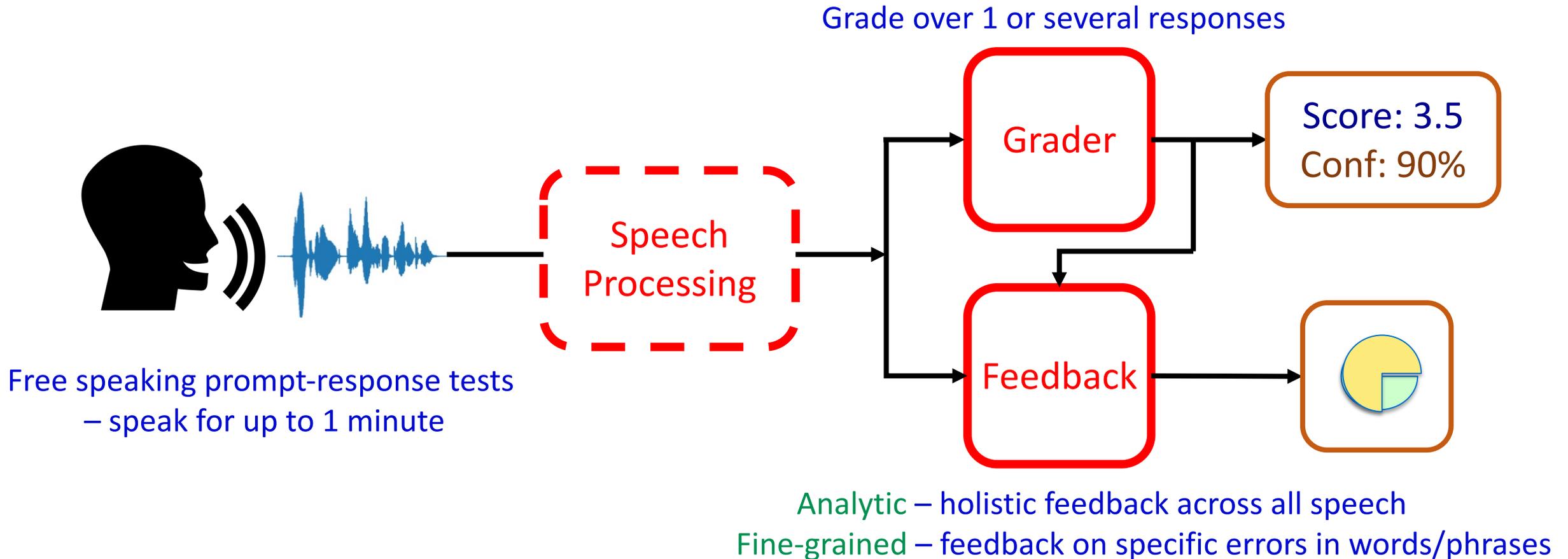
Information encoded in how we speak  
not just what we say

Hesitations

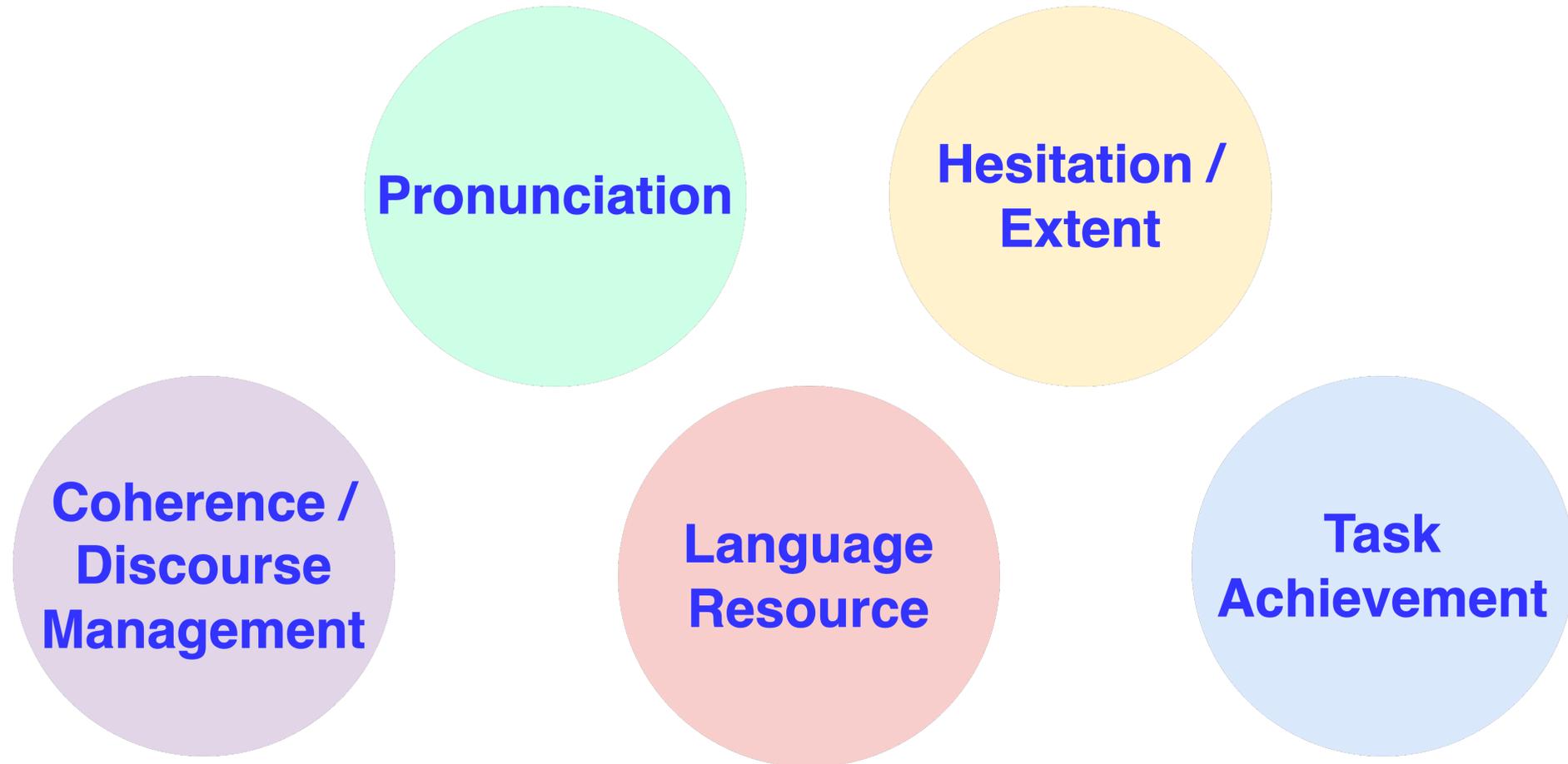
Disfluencies

GENERALY IN JANUARY I ATTENDED IT PROJECT MANAGEMENT TRAINING /.%HESITATION% BECAUSE I ATTEND THE CO  
%HESITATION% TRAINING BECAU DED ME AGAIN TO ATTEND THE  
CO %HESITATION% TRAINING CO ONE VEHICLES ONE-DAY COURSE  
/ IN THE MORNING WE HAVE SIM SKILL TO MANAGE IT  
DEVELOPMENT PROJECT %HESIT THE VENDORS OF CONSULTANTS  
IMPROVED INVOLVED IN THE IT DEVELOPMENT PROJECT IN THEIR PERSON BECAUSE WE HAVE THE LITTLE SESSION HOW  
TO MANAGE A DIFFERENT AS OF CONSULTANTS %HESITATION% TO TO IMPROVE %HESITATION% TO MANAGE THE %HESITATION% IT

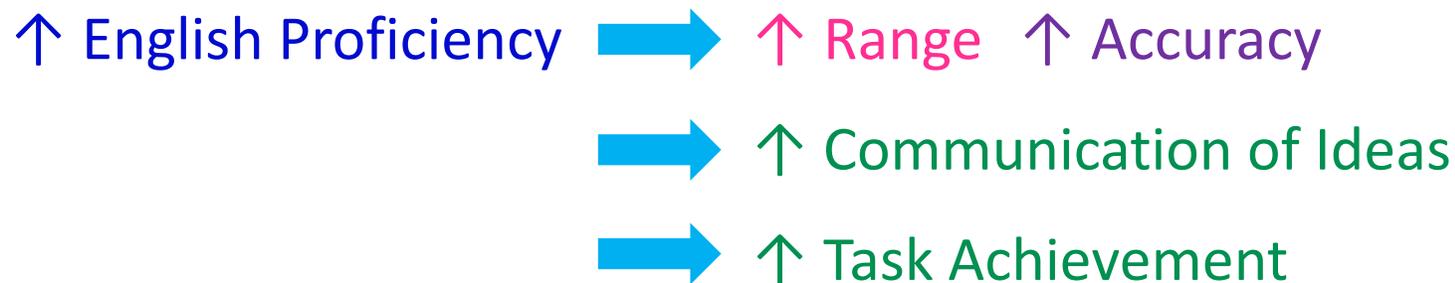
# Spoken Language Assessment and Feedback Pipeline



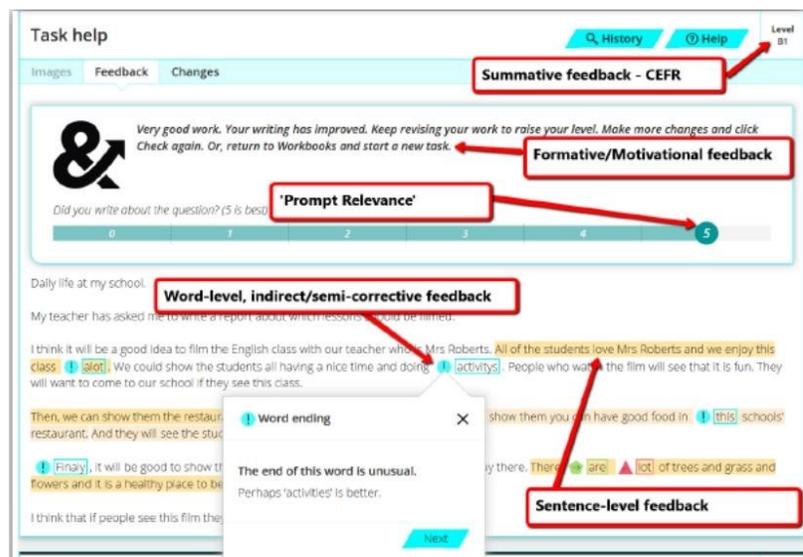
# Speaking core skills



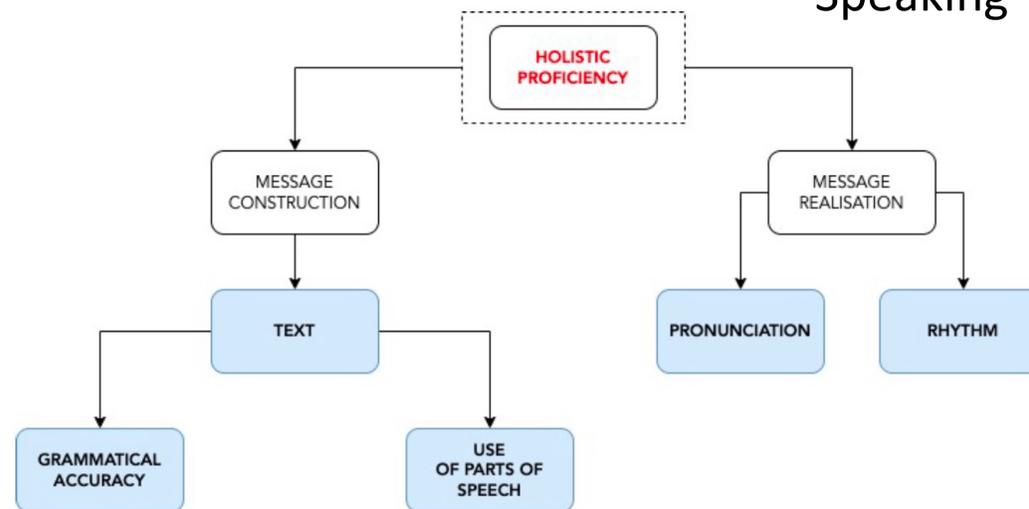
# Language Resource: Grammar and Vocabulary



Writing<sup>1</sup>



Speaking<sup>2</sup>



# Spoken Grammatical Error Correction

# Grammatical Error Correction (GEC)

- Aim of GEC is to produce the grammatically correct sentence from a sentence with mistakes

Original:      The cat **sit** on the mat.

Corrected:     The cat **sat** on the mat.

                  The cat **is sitting** on the mat.



- 1:1 output assumption has relatively small effect on use of GEC
- Large pre-trained GEC models available for English L1 text
  - Lots of annotated L2 learner English writing for fine-tuning
  - **But can we apply these to speech?**

# Challenges and advantages of spoken language

- Challenges for assessing and giving feedback on speech “grammatical” errors
  - We don’t speak in sentences
  - We repeat ourselves, hesitate, mumble etc
  - No defined spoken grammar ... however there are phrases that a L1 speaker is highly unlikely to say
- Advantages of speech
  - No spelling or punctuation mistakes: over 25% written L2 English grammatical errors
  - We provide additional information about meaning of the text within the audio signal (and gestures)

Text + Pronunciation + Prosody (+Delivery)

➤ Spoken Text ≠ Written Text

# Spoken Grammatical Error Correction (SGEC)

Original speech: the cat si- sit on the um mat

Corrected: The cat sat on the mat.



# Challenges for automatic Spoken GEC

- ASR errors are likely

the cat **si-** **sit** on the **um** mat → the cat **si-** **see** on the **um** mat → The cat **saw** on the mat.

- Increased for learner speech as harder to recognise
- Have to automatically segment data into “sentences”
- Lack of training data
  - Very little annotated speech data – insufficient to train on
    - NICT-JLE<sup>1</sup> - no audio samples
    - Linguaskill Speaking<sup>2</sup> (internal data set) – approx. 3000 annotated responses
  - Written text data is mismatched to speaking e.g. no disfluencies, stricter grammar, different style

# Requirements for Spoken GEC for Assessment and Feedback

- Accurate
  - Emphasis on precision with reasonable recall
  - Feedback only on actual grammatical errors – not errors made by the ASR system
    - Giving feedback on an ASR error is worse than giving no feedback
- Interpretable/explainable
  - Let the learner know why what they did was incorrect
- Reasonable computational load and time
  - Near real-time needed for practice apps

# Systems for Spoken GEC



# Spoken GEC – End2end?



| Corpus                              | Audio | Text | DSF | GEC | L2? |
|-------------------------------------|-------|------|-----|-----|-----|
| ASR-Train <sup>1</sup>              | ✓     | ✓    |     |     | ✓   |
| Switchboard <sup>2</sup>            | ✓     | ✓    | ✓   |     |     |
| CLC <sup>3</sup> + BEA <sup>4</sup> |       | ✓    |     | ✓   | ✓   |

**E2E not feasible (currently)**

- No paired training data
- Hard to give feedback to learners

# Spoken GEC – Cascade pipeline



- Train/fine-tune each module on different data sets
- Various sources of information for feedback

# Spoken GEC – Cascade pipeline challenges



➤ Error propagation

➤ Domain mismatch



➤ Prosody information loss



# Automatic Speech Recognition (ASR)

- Hybrid ASR<sup>1</sup>
  - Acoustic model:
    - Kaldi LF-MMI TDNN-F<sup>2</sup> on L2 learner speech (ASR-Train) and AMI
  - Language model:
    - Trigram LM trained on ASR L2 manual transcriptions and Broadcast News English
    - su-RNNLM<sup>3</sup> trained on semi-supervised ASR L2 transcriptions
- Key features:
  - L2 speech from over 25 L1s
  - Explicitly model partial words and map all hesitations to a single word label
  - Word level confidence scores

# Spoken Disfluencies

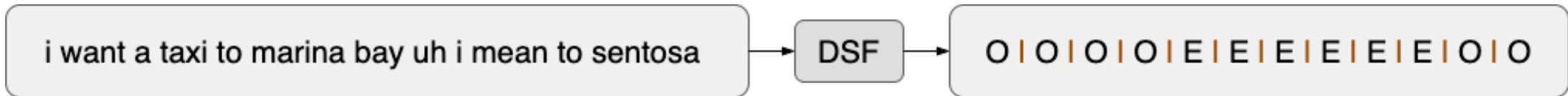
- Typical disfluency

*reperandum*    *interregnum*    *repair*  
I want a taxi [ to Marina Bay uh I mean + to Sentosa ]  
↓  
I want a taxi to Sentosa

- *Interregnum* regions – easy to detect using rule-based methods
- *Reparandum* regions → focus of automatic disfluency detection

# Automatic Disfluency Detection (DD)

- Sequence tagging models



$$\hat{d}_{1:L} = \operatorname{argmax}_{d_{1:L}} P(d_{1:L} | w_{1:L}; \theta_{\text{DD}})$$

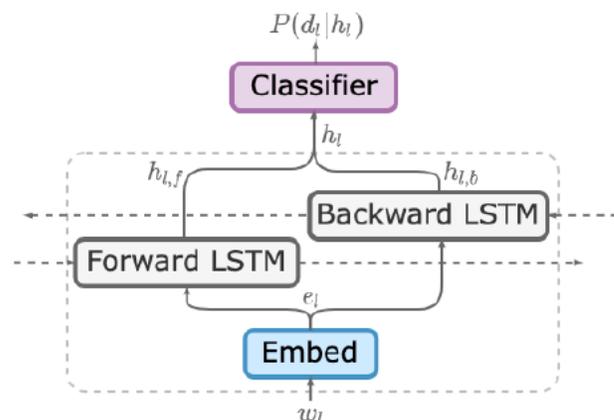
- Non auto-regressive tagger

$$P(d_{1:L} | w_{1:L}; \theta_{\text{DD}}) = \prod_{l=1}^L P(d_l | d_{1:l-1}, w_{1:L}; \theta_{\text{DD}}) \approx \prod_{l=1}^L P(d_l | w_{1:L}; \theta_{\text{DD}})$$

- Performance largely dependent on quality of tagger's internal feature representations
  - Depends on training criteria and framework<sup>1</sup>

# Sequence Tagging DD Models

- Recurrent neural network (RNN)<sup>1,2</sup>



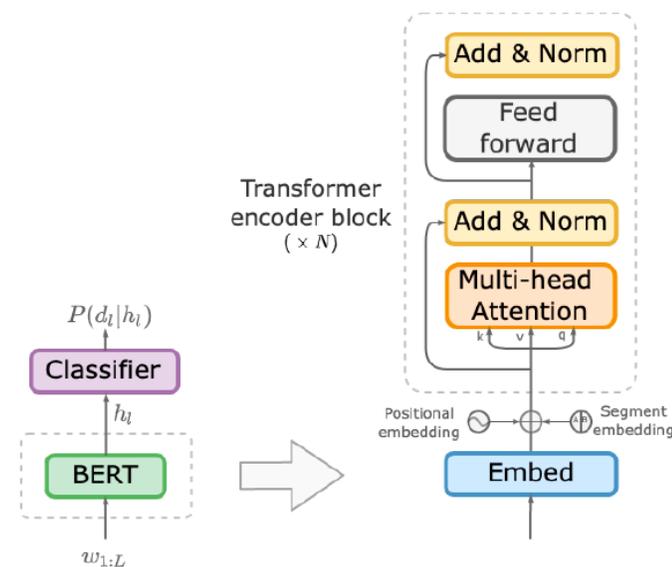
$$e_l = \mathbf{W}^e w_l$$

$$\vec{h}_l = \text{LSTM}(e_l, \vec{h}_{l-1}; \vec{\theta}) \quad \overleftarrow{h}_l = \text{LSTM}(e_l, \overleftarrow{h}_{l+1}; \overleftarrow{\theta})$$

$$h_l = [\vec{h}_l, \overleftarrow{h}_l]$$

- Reasonable good quality feature reps
- Training scales with sequence lengths

- Large pre-trained model: BERT<sup>3</sup>



$$h_{1:L} = \text{BERT}(w_{1:L})$$

- Capable of high quality feature reps
- Fine-tune BERT using DD objective



$$\hat{y}_{1:N} = \operatorname{argmax}_{y_{1:N} \in \mathcal{Y}} P(y_{1:N} | w_{1:L}; \theta_{\text{GEC}})$$

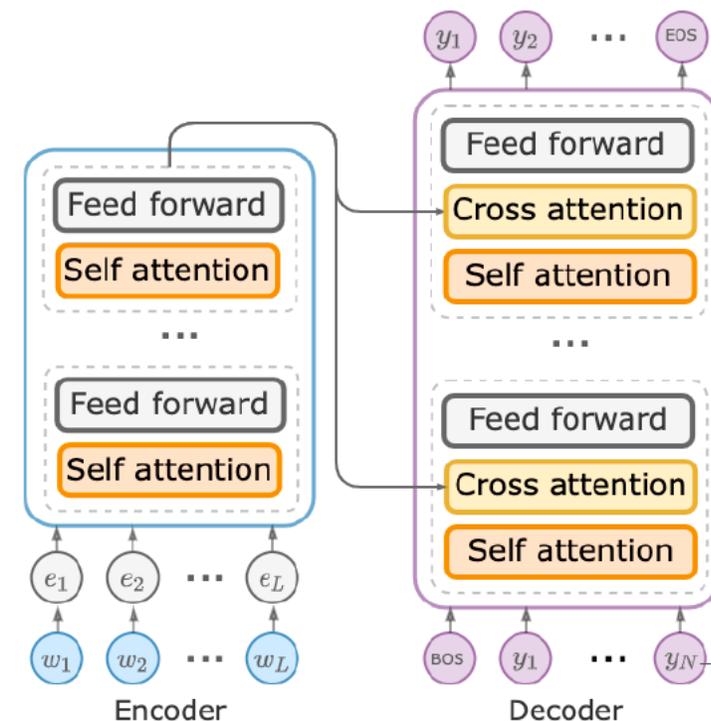
# seq2seq GEC

Treat GEC as a machine translation problem i.e. seq2seq

- Encoder-decoder architecture
- RNN-based → Transformer-based encoder-decoder models
  - Large pretrained e.g. T5 'Text-to-Text Transfer Transformer'<sup>1</sup>
- **Gramformer<sup>2</sup>**: T5-based fine-tuned for GEC on various corpora
  - Can further fine-tune to target domain GEC
  - 223M parameters



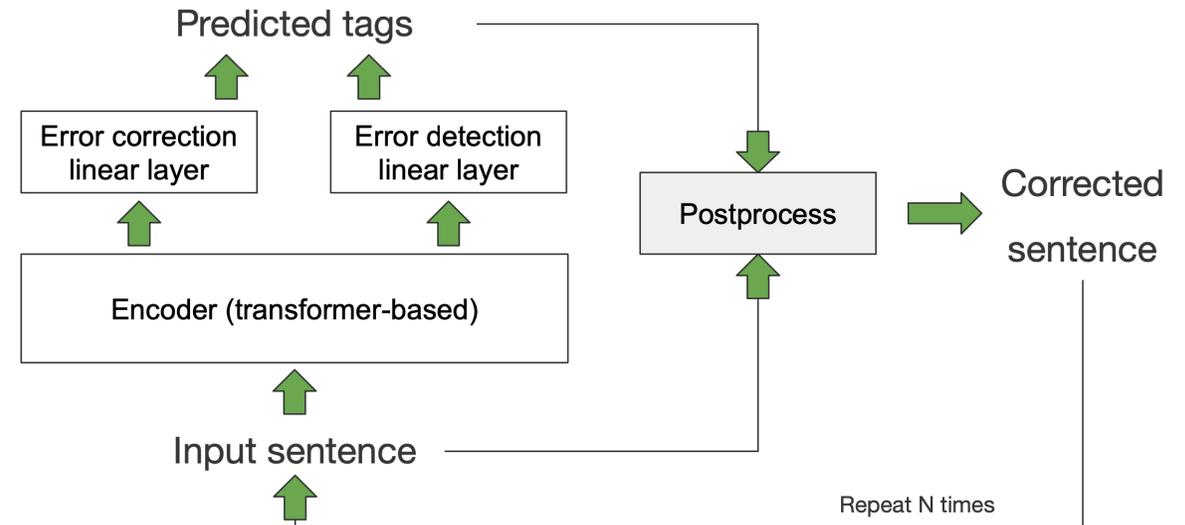
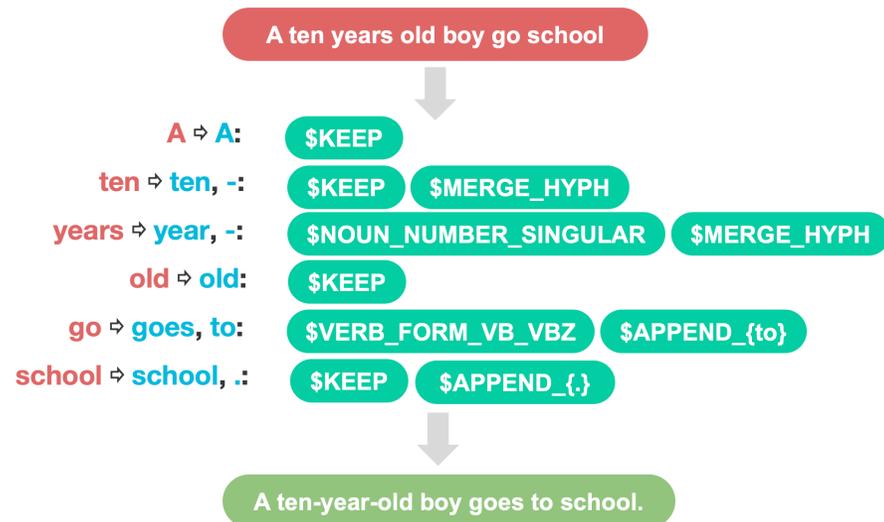
- **ChatGPT<sup>3</sup>**: generative autoregressive large language model [CGPT]
  - 175B parameters



# Sequence tagging GEC

Treat GEC as a sequence tagging problem<sup>1,2</sup>

- For each token in the source sequence GEC should produce a tag (edit) for required correction operation
  - e.g. GECToR by Grammarly (138M parameters)



GECToR architecture (from O. Skurzhanyski and K. Omelianchuk<sup>3</sup>)

# SGEC System Corpora

| Corpus              | Spoken? | #sents | #words | % disfluencies |
|---------------------|---------|--------|--------|----------------|
| <b>Training</b>     |         |        |        |                |
| CLC <sup>1</sup>    | X       | 1.9M   | 25.2M  | -              |
| BEA <sup>2</sup>    | X       | 1M     | 11.5M  | -              |
| <b>Evaluation</b>   |         |        |        |                |
| FCEtst <sup>3</sup> | X       | 2681   | 37k    | -              |
| LIN <sup>4</sup>    | ✓       | 3361   | 38k    | 5%             |



1. D. Nicholls, "The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT", Corpus Linguistics 2003 conference

2. C. Bryant et al, "The BEA-2019 shared task on grammatical error correction". BEA Workshop 2019

3. H. Yannakoudakis et al, "A new dataset and method for automatically grading ESOL texts", NAACL-HLT 2011

4. Y. Lu et al, "On Assessing and Developing Spoken 'Grammatical Error Correction' Systems", BEA Workshop 2022

# SGEC training data pre-processing

- Convert written text corpora for training to “speechify” it
  - Correct any spelling errors
  - Remove punctuation
  - Upper (or lower) case all text to match ASR
  - Adding speech-like disfluencies through a masked LM can help (not used here)

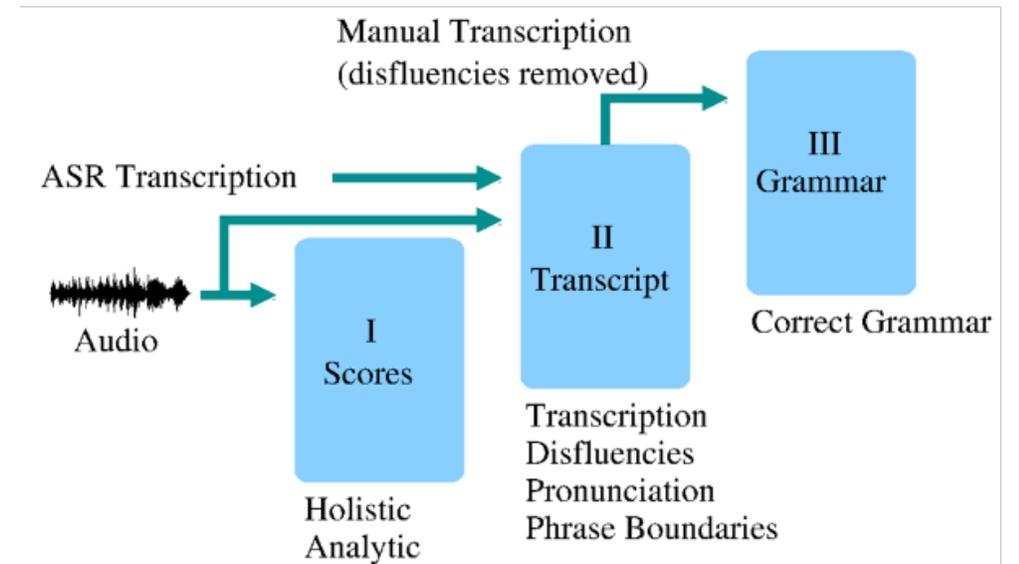
It was a cleer, blue, sky.



IT WAS A CLEAR BLUE SKY

# SGEC spoken evaluation data

- LIN: Linguaskill Speaking L2 learner speech from multi-level English test
  - 3361 grammatically corrected 'sentences' from 972 question responses
  - 340 speakers from over 30 L1s
  - Approximately equal distribution over grades A1-C1
- ASR and Manual transcriptions available
  - With/out disfluencies
  - Manually annotated by English Language iTutoring (ELiT) Ltd



# Phase 2 annotation

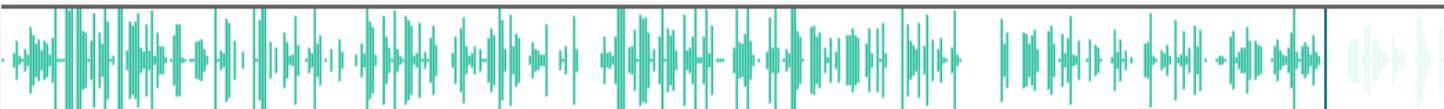
## ^ Previous Annotations

## ∨ Answer



Long turn 1

Talk about a training course you attended for your work. You should say: • what the course was about • why you went on the course • what you learnt from it.



⏪ ⏩ | ... ? f P | / /? /- | 👁 Original **Annotated** Corrected | 🗑 Disfluency Pronunciation Partial Backchannel

GENERALLY IN JANUARY I ATTENDED **FF** **L-I** PROJECT MANAGEMENT TRAINING **/.** **%HESITATION%** **BECAUSE** I ATTEND THE **CO**  
**%HESITATION%** TRAINING BECAUSE **M** MY BOSS **%HESITATION%** DIRECTOR **KINDA** **GENERAL** RECOMMENDED ME **AGAIN** **TO ATTEND** THE  
**CO** **%HESITATION%** TRAINING COURSE **/.** THE TRAINING COURSE **COURSES** **CONSISTS OF** TWO PARTS **IN** **ONE VEHICLES** **ONE-DAY** COURSE  
**/.** IN THE MORNING WE HAVE **SIMILAR** **SEMINAR** TO LEARN ABOUT **%HESITATION%** **%HESITATION%** BASIC SKILL TO MANAGE **FF** **L-I**  
DEVELOPMENT PROJECT **%HESITATION%** FOR EXAMPLE **%HESITATION%** **TWO TO** **MANDATES** **MANAGE THE** **VENDORS** **?** **OF** CONSULTANTS  
**IMPROVED** **INVOLVED** IN THE **L-I** DEVELOPMENT PROJECT **/.** IN THE AFTERNOON **BECAUSE** **COURSE** WE HAVE THE **LITTLE** **?** SESSION HOW  
TO MANAGE A **DIFFERENT-AS** **?** **OF** CONSULTANTS **%HESITATION%** **TO TO** **IMPROVE** **%HESITATION%** TO MANAGE **THE** **%HESITATION%** **FF** **L-I**  
**/-**

# Annotation phase 2 to phase 3

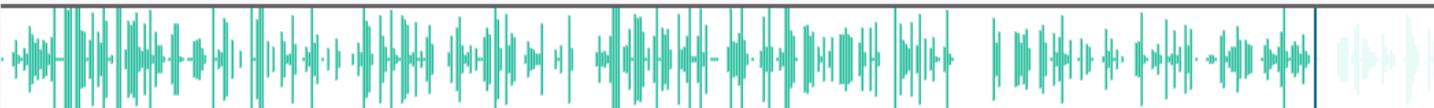
## ^ Previous Annotations

### ∨ Answer



Long turn 1

Talk about a training course you attended for your work. You should say: • what the course was about • why you went on the course • what you learnt from it.



⏪ ⏩ | ... ? f P | / /? /- | 👁 Original Annotated **Corrected** | 🗑 Disfluency Pronunciation Partial Backchannel

IN JANUARY I ATTENDED I-T PROJECT MANAGEMENT TRAINING. I ATTEND THE TRAINING BECAUSE MY BOSS DIRECTOR GENERAL RECOMMENDED ME TO ATTEND THE TRAINING COURSE. THE TRAINING COURSE CONSISTS OF TWO PARTS IN ONE-DAY COURSE. IN THE MORNING WE HAVE SEMINAR TO LEARN ABOUT BASIC SKILL TO MANAGE I-T DEVELOPMENT PROJECT FOR EXAMPLE TO MANAGE THE OF CONSULTANTS INVOLVED IN THE I-T DEVELOPMENT PROJECT. IN THE AFTERNOON COURSE WE HAVE THE SESSION HOW TO MANAGE A OF CONSULTANTS TO MANAGE THE I-T.

# Annotation phase 3

## ▼ Question

Long turn 1

Talk about a piece of technology you would like to own. You should say: • which piece of technology you would like to have • why you would like this technology • whether you think you will have this technology one day.



Original

Annotated

Corrected



Spelling

THE PIECE OF TECHNOLOGY THAT I WOULD LIKE TO HAVE IS A HYBRID CAR. I THINK THAT CURRENTLY WITH CONCERNS FOR ABOUT THE ENVIRONMENT WHICH CAN COULD BE IN A CATASTROPHIC STATE IF WE DON'T TAKE ACTION SOON ENOUGH HAVING A HYBRID CAR WOULD BE AN INCREDIBLE ADVANTAGE. I THINK IT'S RATHER QUITE REALISTIC THAT I WOULD WILL HAVE THIS TECHNOLOGY ONE DAY PROVIDED THAT THE PRICE ISN'T TOO STEEP AND THAT CERTAIN PUBLIC ACCOMMODATIONS FACILITIES SUCH AS FOR RECHARGING BATTERY LIFE WOULD BE ACCESSIBLE TO THE CURRENT GENERAL PUBLIC.

- Minimal edit corrections made
- Type of error not marked

# Measuring SGEN performance

- Align and classify GEC edits with ERRor ANnotation Toolkit (ERRANT)<sup>1,2</sup>
  - Automatically extracts edits from parallel original and corrected sentences
  - Classifies them according to a dataset-agnostic rule-based framework
  - Facilitates error type evaluation at different levels of granularity

|            |     |     |        |    |       |     |
|------------|-----|-----|--------|----|-------|-----|
| Auto:      | the | cat | sit    | on |       | mat |
| Reference: | the | cat | sat    | on | the   | mat |
| Edit:      |     |     | R:VERB |    | M:DET |     |

- Spoken language differences to written text
  - Alignment will change depending on if Manual or ASR reference
  - Edit classification rules do not take speech effects into account such as disfluencies and partial phrases

# Spoken GEC Performance

- Manual, Fluent speech transcripts → Gramformer

|                        | <b>FCEtst</b> | <b>LIN</b> |
|------------------------|---------------|------------|
| $M^2 F_{0.5} \uparrow$ | 56.6          | 53.6       |

- Spoken GEC can match performance of Written GEC L2 English learner task

- Cascade GEC system: TDNN-F ASR → BERT DSF → Gramformer

|                     | <b>LIN</b>    |                   |
|---------------------|---------------|-------------------|
| Input to Gramformer | <b>Manual</b> | <b>ASR+DD+GEC</b> |
| SER ↓               | 43.3          | 76.8              |
| TER ↓               | 8.3           | 27.9              |

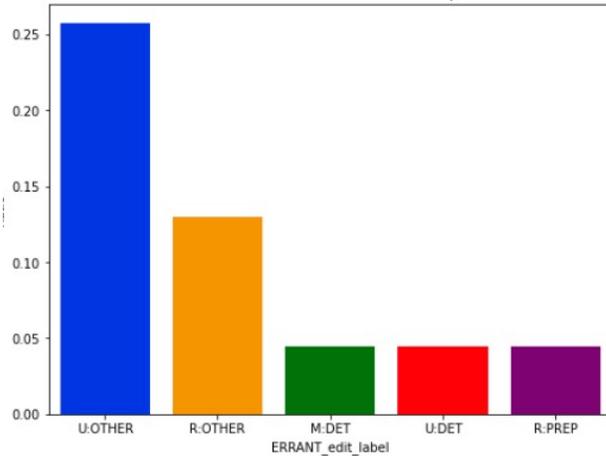
- Need to mitigate for ASR error propagation and domain mismatch

# GEC system analysis



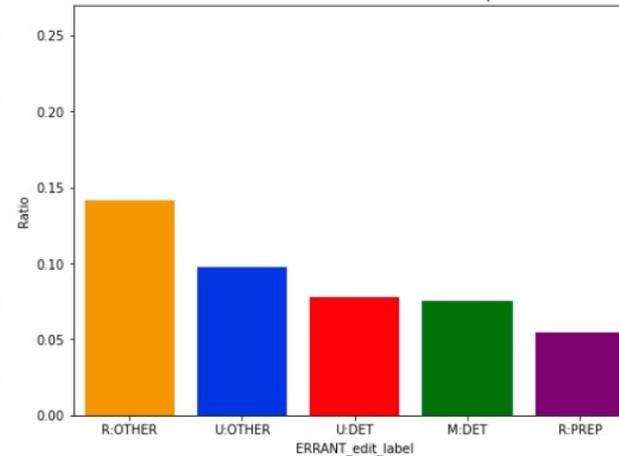
|   | ERRANT_edit_label | Abs_count | Ratio    |
|---|-------------------|-----------|----------|
| 0 | U:OTHER           | 1355      | 0.257507 |
| 1 | R:OTHER           | 682       | 0.129609 |
| 2 | M:DET             | 235       | 0.044660 |
| 3 | U:DET             | 234       | 0.044470 |
| 4 | R:PREP            | 233       | 0.044280 |

5 most common ERRANT edit labels - Manual transcriptions vs manual GEC



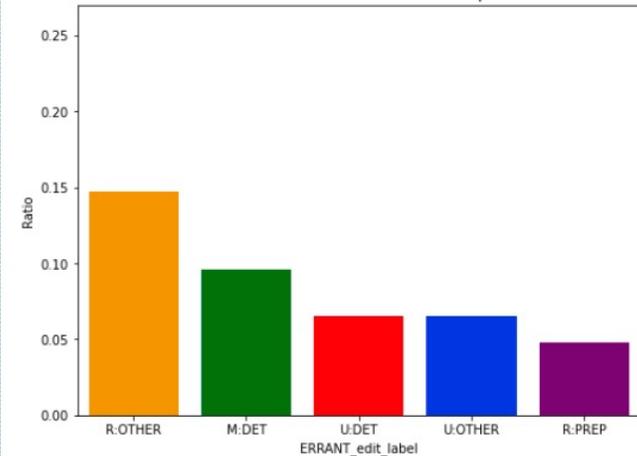
|   | ERRANT_edit_label | Abs_count | Ratio    |
|---|-------------------|-----------|----------|
| 0 | R:OTHER           | 389       | 0.141403 |
| 1 | U:OTHER           | 269       | 0.097783 |
| 2 | U:DET             | 214       | 0.077790 |
| 3 | M:DET             | 206       | 0.074882 |
| 4 | R:PREP            | 151       | 0.054889 |

5 most common ERRANT edit labels - Manual transcriptions vs auto GEC



|   | ERRANT_edit_label | Abs_count | Ratio    |
|---|-------------------|-----------|----------|
| 0 | R:OTHER           | 439       | 0.147662 |
| 1 | M:DET             | 284       | 0.095526 |
| 2 | U:DET             | 194       | 0.065254 |
| 3 | U:OTHER           | 193       | 0.064918 |
| 4 | R:PREP            | 142       | 0.047763 |

5 most common ERRANT edit labels - ASR transcriptions vs auto GEC

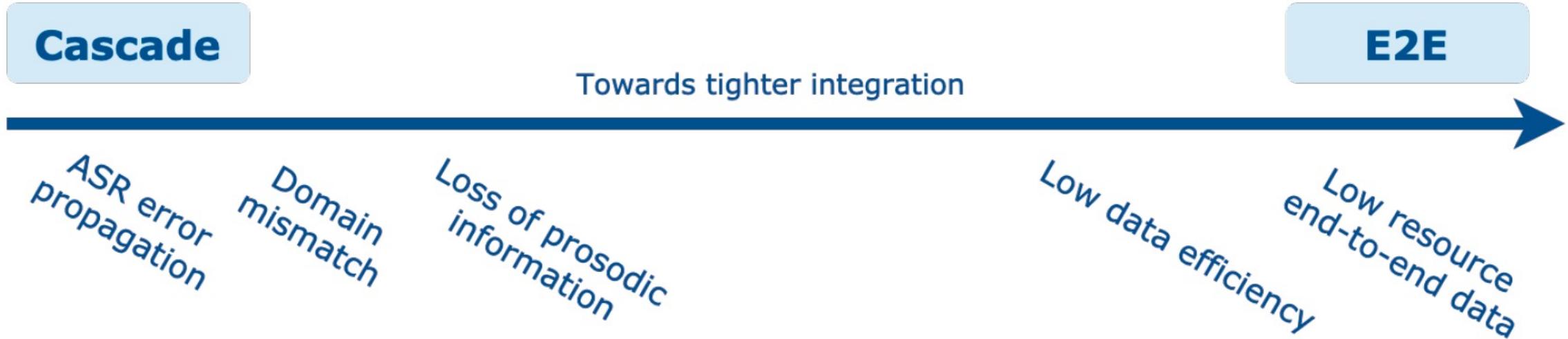


- Over prediction of OTHER due to ERRANT missing speech-based rules

\* This system did not have DD module



# Integrated Spoken GEC Training

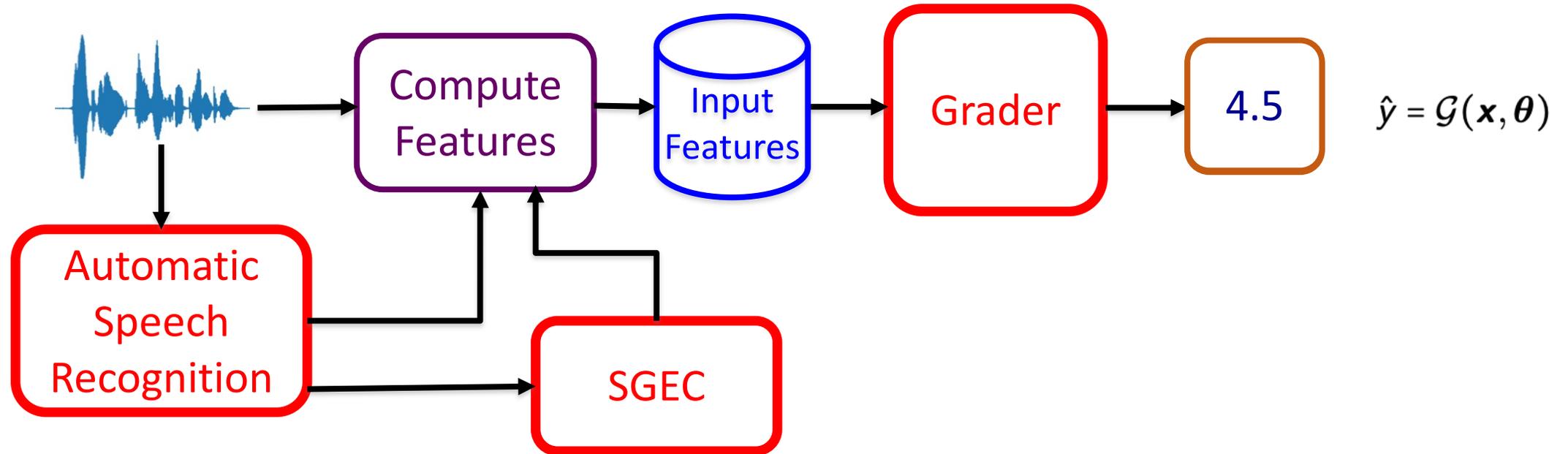


- Approaches towards tighter integration
  - Semi-supervised ASR error mitigation
  - Reranking
  - Embedding passing

# Using Spoken GEC in Assessment

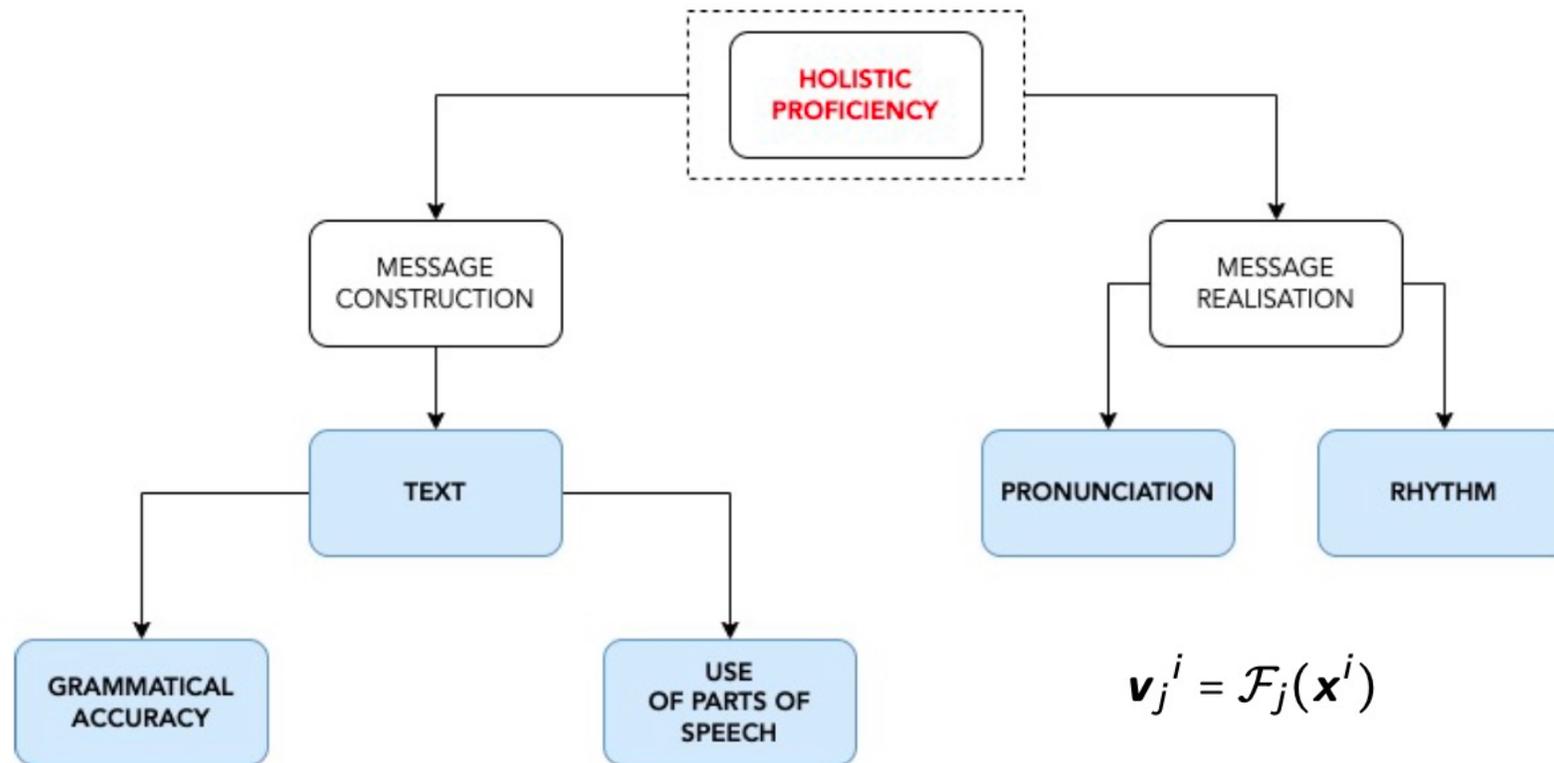


# Holistic “hand-crafted” features-based auto-marker



Standard approach to speech auto-marking

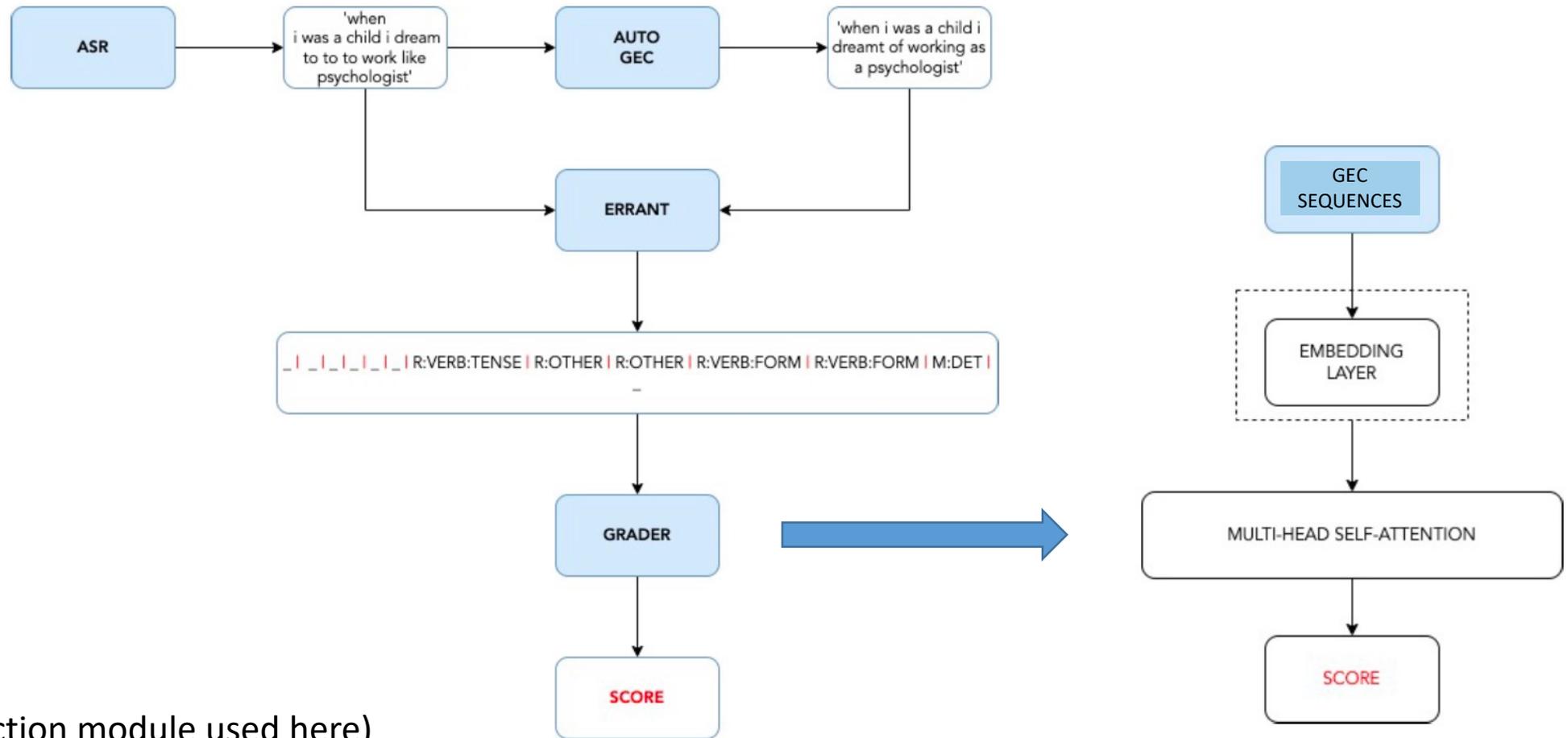
# View Specific and/or Multi-view auto-marker<sup>1</sup>



$$\mathbf{v}_j^i = \mathcal{F}_j(\mathbf{x}^i)$$

$$\hat{y}_j^i = \mathcal{G}_j(\mathbf{v}_j^i, \theta)$$

# Grammatical accuracy grader pipeline



(no disfluency detection module used here)

# GEC Grader evaluation

- ERRANT label distribution over grades



|    | A1    | A2    | B1    | B2    | C     |
|----|-------|-------|-------|-------|-------|
| A1 | 0.0   | 0.067 | 0.362 | 0.808 | 1.307 |
| A2 | 0.067 | 0.0   | 0.139 | 0.472 | 0.869 |
| B1 | 0.362 | 0.139 | 0.0   | 0.103 | 0.327 |
| B2 | 0.808 | 0.472 | 0.103 | 0.0   | 0.074 |
| C  | 1.307 | 0.869 | 0.327 | 0.074 | 0.0   |

- Grader RMSE

| Model            | LinBus       |
|------------------|--------------|
| holistic         | 0.406        |
| <b>GEC</b>       | <b>0.495</b> |
| text             | 0.409        |
| pronunciation    | 0.451        |
| 5-way multi-view | 0.386        |

# Using Spoken GEC for Learner/Teacher Feedback

# What and how to present?

### The Feedback Generation Pipeline for Speech Correction

these **is** **are** my dogs

| Incorrect Word(s) | Corrected | Error made                        | Error Description  |
|-------------------|-----------|-----------------------------------|--|
| is                | are       | Replacment Subject-Verb Agreement | False verb form for the given subject. e.g. (He) have+(He) has |

#### Past Errors Made

that **cause** **causes** micky to left em

in doing so minni **is loosing** **looses** some notes to the song turqey in the straw as well as a small guitar

takes **the** **a** boat

i should have **went** **gone** somewhere

A. Filighera et al, "Towards A Vocalization Feedback Pipeline for Language Learners," 2022 ICALT

Report details (报告详情)

Verb tenses

You said: [Incorrect icon]

You should: [Correct icon]

There were a few mistakes in terms of your verb tenses. Please practice more and be careful about how you use verbs next time.

Present simple third-person singular verb '-s'

You said: [Incorrect icon] You should: [Correct icon]

**YOU SHOULD SAY**

There were a few mistakes in terms of your verb tenses. Please practice more and be careful about how you use verbs next time.

<https://liulishuo.com>

Task help

Summative feedback - CEFR

Formative/Motivational feedback

Prompt Relevance

Word-level, indirect/semi-corrective feedback

Sentence-level feedback

Word ending: The end of this word is unusual. Perhaps 'activities' is better.

<https://writeandimprove.com>

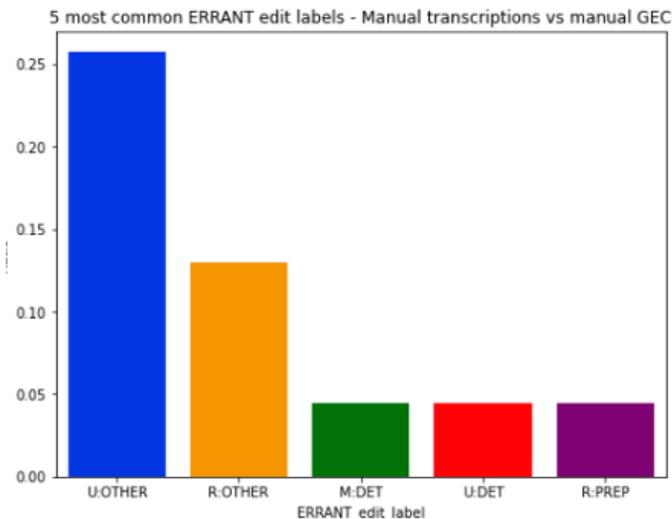
# Improve feedback precision: remove 'OTHER'

ERRANT output:

```
S we share a lot of patience and there are lot of opinion income
A 5 6| | R:NOUN| |opinions| |REQUIRED| | -NONE-| | | 0
A 9 9| | M:DET| |a| |REQUIRED| | -NONE-| | | 0
A 11 13| | R:OTHER| | things in common| |REQUIRED| | -NONE-| | | 0
```

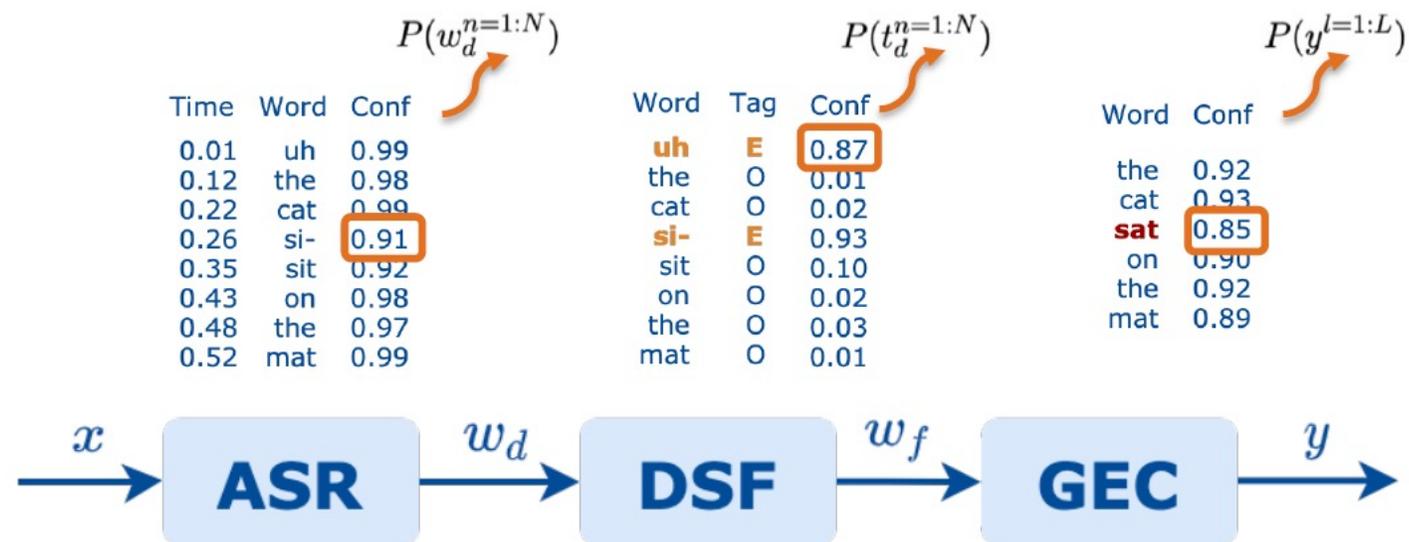
Position

Type Correction



- High number of OTHER tags in manual → automatic labels
  - Ambiguous type, not informative to learners
  - May be a paraphrase
  - Don't present to learner

# Improve feedback precision: confidence filtering



$$\log P = \alpha \log P_{\text{ASR}} + \beta \log P_{\text{DSF}} + \gamma \log P_{\text{GEC}}$$

- Sentence level > edit level confidence filtering

# Confidence filtering type analysis

| Edit Type    | ELIT-ASR-AUTO |              |
|--------------|---------------|--------------|
|              | None          | Sent         |
| M:DET        | 19.12         | <b>27.42</b> |
| R:PREP       | 21.54         | <b>33.33</b> |
| R:NOUN:NUM   | 22.58         | <b>38.39</b> |
| R:VERB:TENSE | <b>24.54</b>  | 21.40        |
| U:DET        | 17.19         | <b>21.85</b> |
| R:VERB       | 22.61         | <b>24.14</b> |
| R:NOUN       | <b>27.87</b>  | 12.46        |
| M:PREP       | 12.15         | <b>18.32</b> |
| R:VERB:FORM  | 22.37         | <b>29.93</b> |
| R:VERB:SVA   | 16.85         | <b>34.55</b> |
| TOTAL        | 20.89         | <b>24.28</b> |

With deterministic correction

**IMPROVE**

With more than 1 feasible correction

**DEGRADE**

Fo.5 breakdown – top 10 edit types

# Open Challenges for Spoken GEC

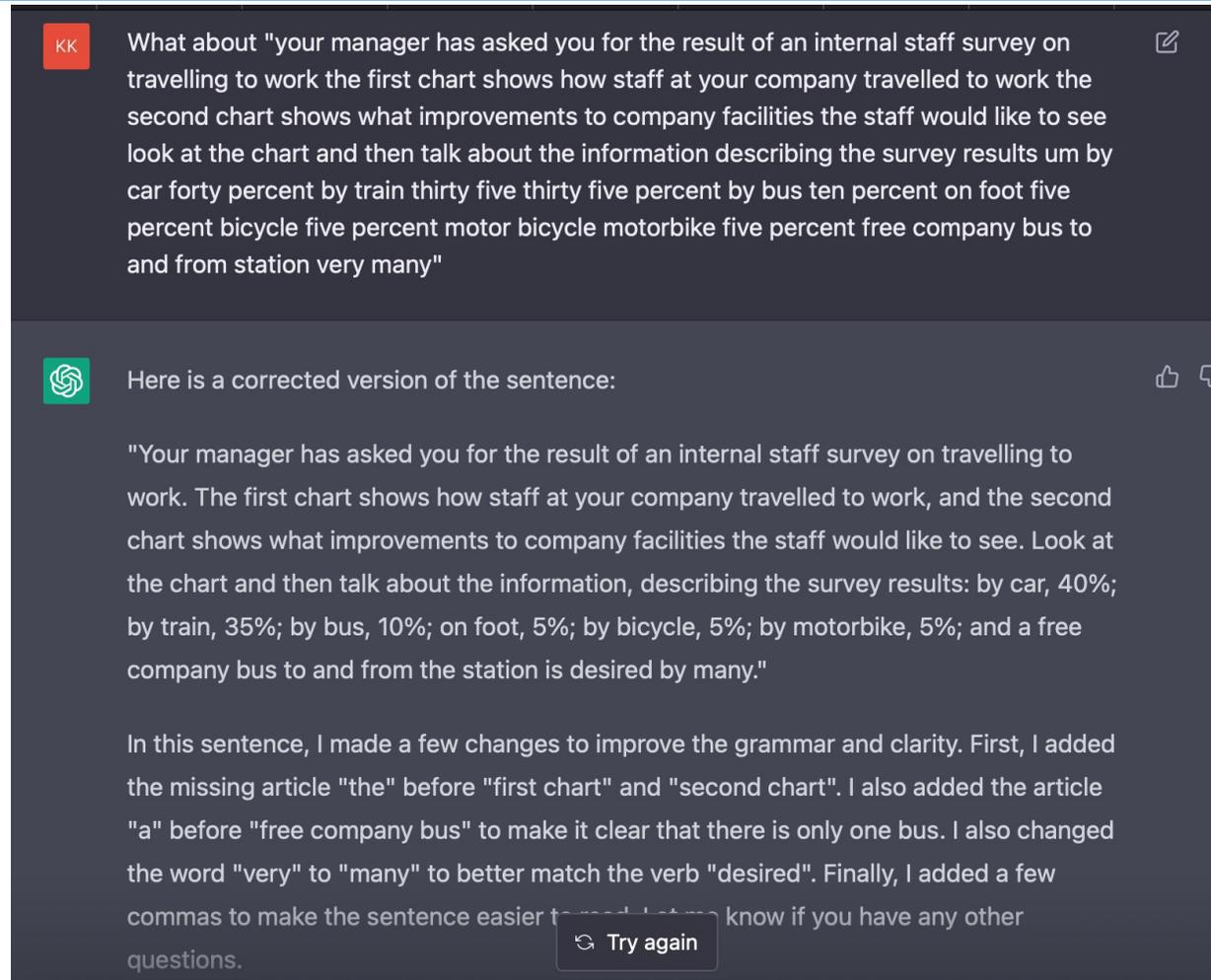
- Increasing Accuracy
  - Tighter integration of modules
  - Data augmentation
  - Domain adaptation of GEC networks
  - Sequence uncertainty – when was a good/bad GEC sequence generated?

- Interpretability/explainability



Saliency: What caused the grammatical error?

# ChatGPT: are huge amounts of data and compute “all we need”?



KK What about "your manager has asked you for the result of an internal staff survey on travelling to work the first chart shows how staff at your company travelled to work the second chart shows what improvements to company facilities the staff would like to see look at the chart and then talk about the information describing the survey results um by car forty percent by train thirty five thirty five percent by bus ten percent on foot five percent bicycle five percent motor bicycle motorbike five percent free company bus to and from station very many" 

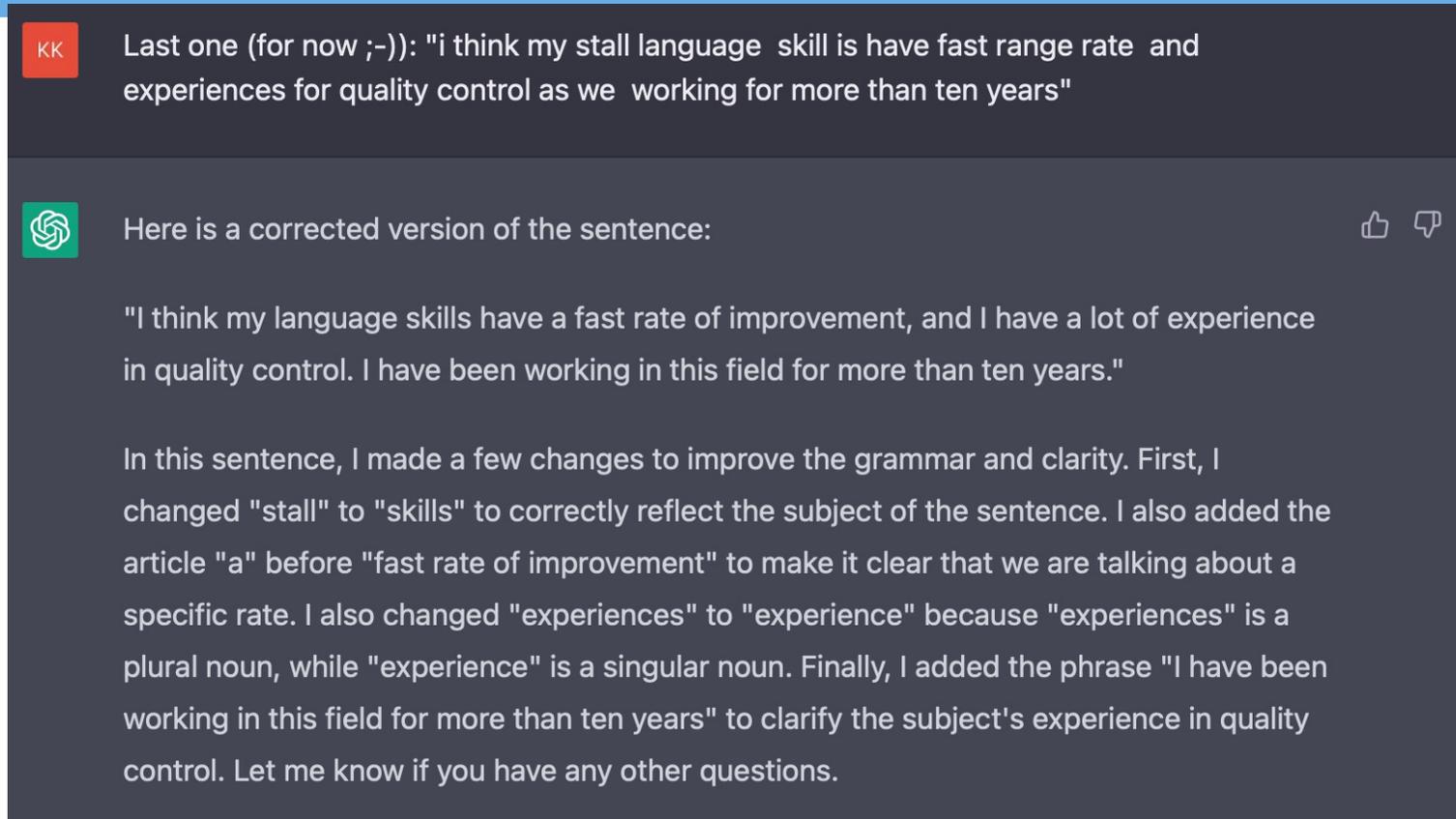
 Here is a corrected version of the sentence:  

"Your manager has asked you for the result of an internal staff survey on travelling to work. The first chart shows how staff at your company travelled to work, and the second chart shows what improvements to company facilities the staff would like to see. Look at the chart and then talk about the information, describing the survey results: by car, 40%; by train, 35%; by bus, 10%; on foot, 5%; by bicycle, 5%; by motorbike, 5%; and a free company bus to and from the station is desired by many."

In this sentence, I made a few changes to improve the grammar and clarity. First, I added the missing article "the" before "first chart" and "second chart". I also added the article "a" before "free company bus" to make it clear that there is only one bus. I also changed the word "very" to "many" to better match the verb "desired". Finally, I added a few commas to make the sentence easier to read. Let me know if you have any other questions.

 Try again

# ChatGPT: ASR errors are still an issue



The screenshot shows a chat interface with a dark background. On the left, a red square with the letters 'KK' in white represents the user's input. The text of the input is: "Last one (for now ;-)): "i think my stall language skill is have fast range rate and experiences for quality control as we working for more than ten years"". On the right, a green square with the OpenAI logo represents the AI's response. The text of the response is: "Here is a corrected version of the sentence: 'I think my language skills have a fast rate of improvement, and I have a lot of experience in quality control. I have been working in this field for more than ten years.'" Below this, the AI provides a detailed explanation of the corrections made to the original sentence.

KK Last one (for now ;-)): "i think my stall language skill is have fast range rate and experiences for quality control as we working for more than ten years"

Here is a corrected version of the sentence:

"I think my language skills have a fast rate of improvement, and I have a lot of experience in quality control. I have been working in this field for more than ten years."

In this sentence, I made a few changes to improve the grammar and clarity. First, I changed "stall" to "skills" to correctly reflect the subject of the sentence. I also added the article "a" before "fast rate of improvement" to make it clear that we are talking about a specific rate. I also changed "experiences" to "experience" because "experiences" is a plural noun, while "experience" is a singular noun. Finally, I added the phrase "I have been working in this field for more than ten years" to clarify the subject's experience in quality control. Let me know if you have any other questions.

- ASR error: “I think my strongest skill...” → “I think my stall language skill ...”

# Conclusions

- Automated assessment and computer aided language learning is growing rapidly
  - Needed to cope with over 2Bn learners/speakers
- Learners increase in grammatical accuracy and range as they progress
  - Automatic Grammatical Error Correction (GEC) is very popular for Writing Assessment and Feedback
  - Automatic Spoken GEC (SGEC) shown to be of use to Speaking Assessment and Feedback
- Lots of open challenges in Automatic SGEC
  - Improving core accuracy whilst handling limited labelled training data
  - Increase robustness to and awareness of ASR errors
  - Add interpretability and explainability to give more in-depth feedback

# Thanks

*Yiting (Edie) Lu and Stefano Bannò for multiple pictures – in addition to their research with Prof Mark Gales and past and present members of the ALTA SLP Technology Project, Cambridge University Engineering Department.*

*Diane Nicholls and the humannotar team at ELiT for the Linguaskill annotations.*

*This presentation reports on research supported by Cambridge University Press & Assessment, a department of The Chancellor, Masters, and Scholars of the University of Cambridge.*

**Questions?**

