



Universidade de Vigo



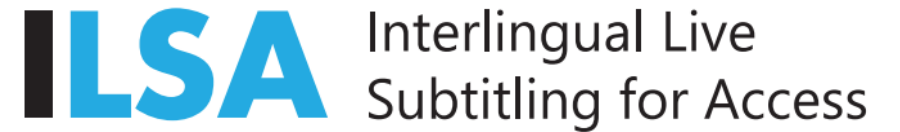
A comparative analysis of interpreters' and subtitlers' performance in interlingual live subtitling

Hayley Dawson, University of Roehampton

Pablo Romero-Fresco, Universidade de Vigo



Aims



Overall aim for ILSA: To design, develop, test and validate the first training course for ILS and provide a protocol for this discipline for TV, the classroom and parliament.

Aim of the main experiment:

To train and test participants in an ILS course to answer the following questions:

- Is ILS feasible?
- Who is better suited?
- What are the main challenges?

The experiment: a short online course

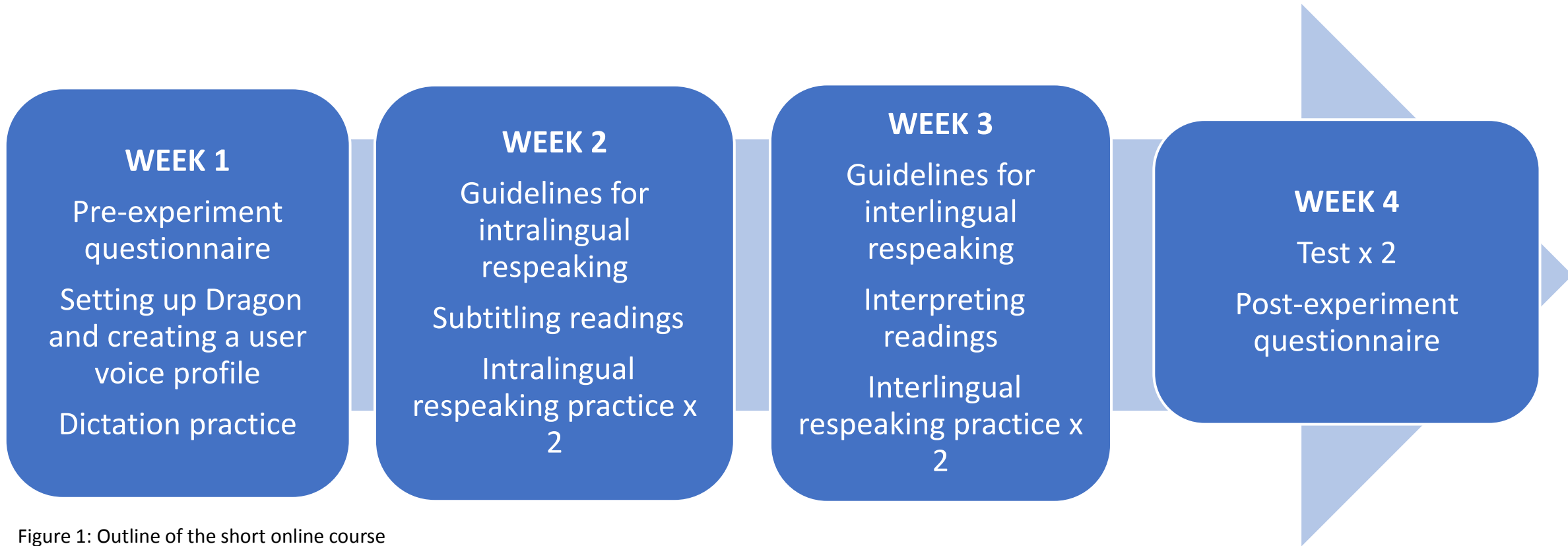


Figure 1: Outline of the short online course

Progress of analysis

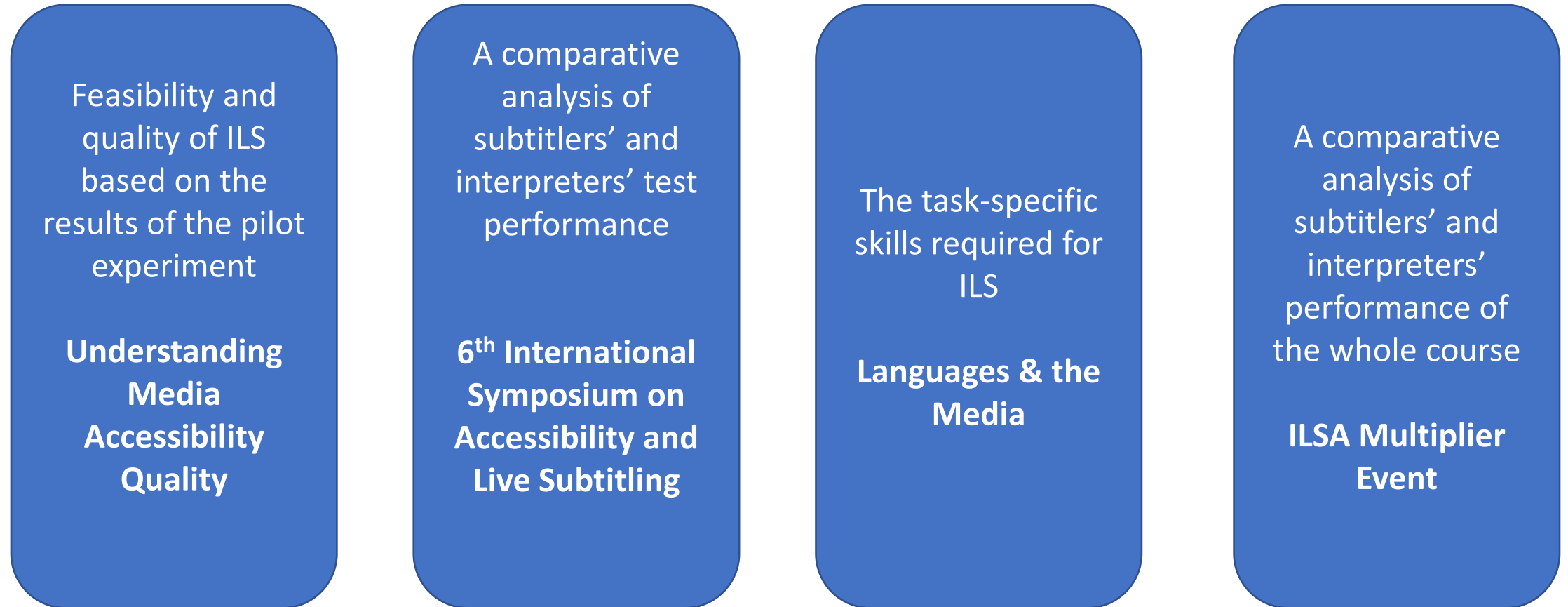


Figure 2: Progress of analysis

Respeaking exercises

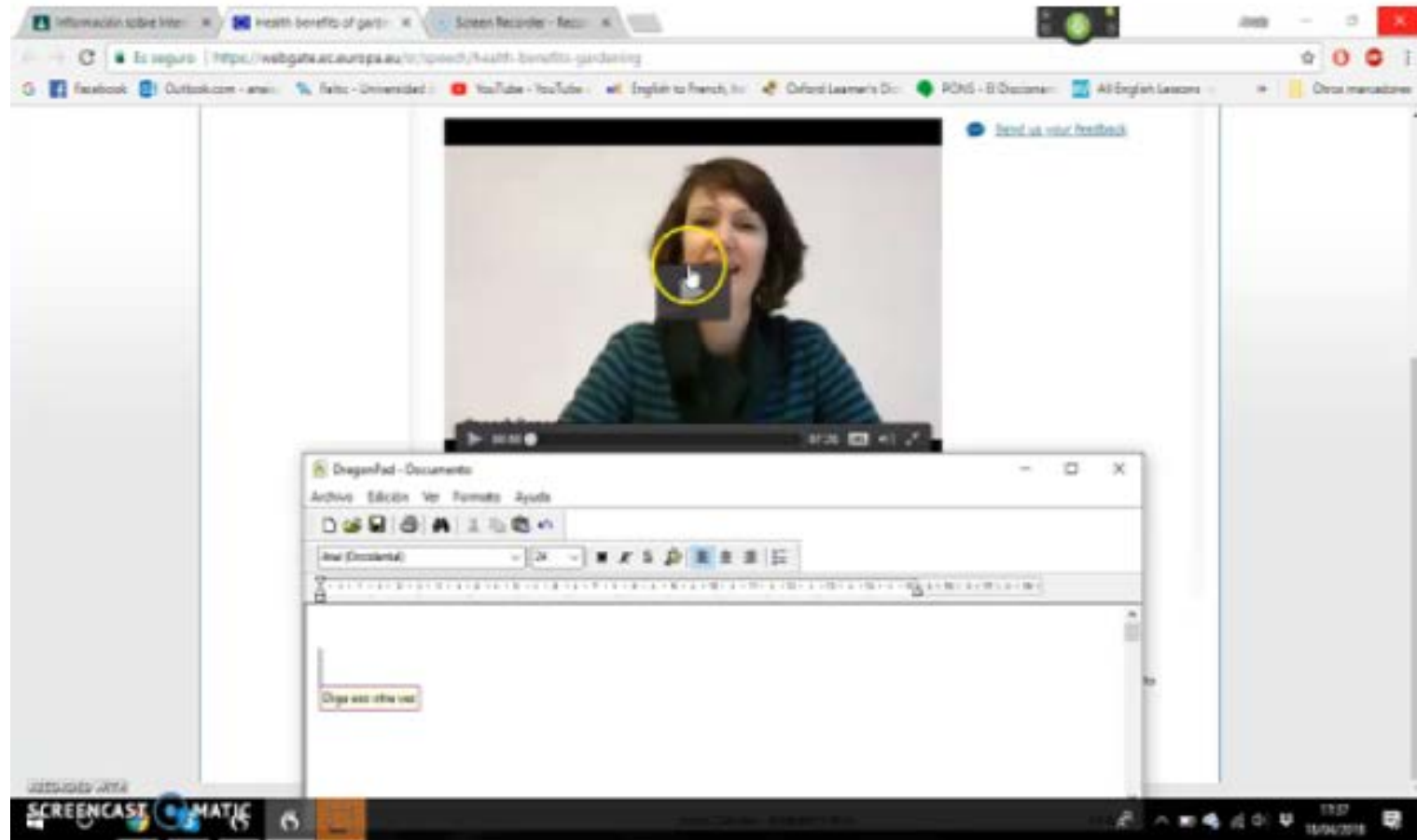


Figure 3: Example of an interlingual respelling test

The NTR model

(Romero-Fresco & Pöchhacker, 2017)

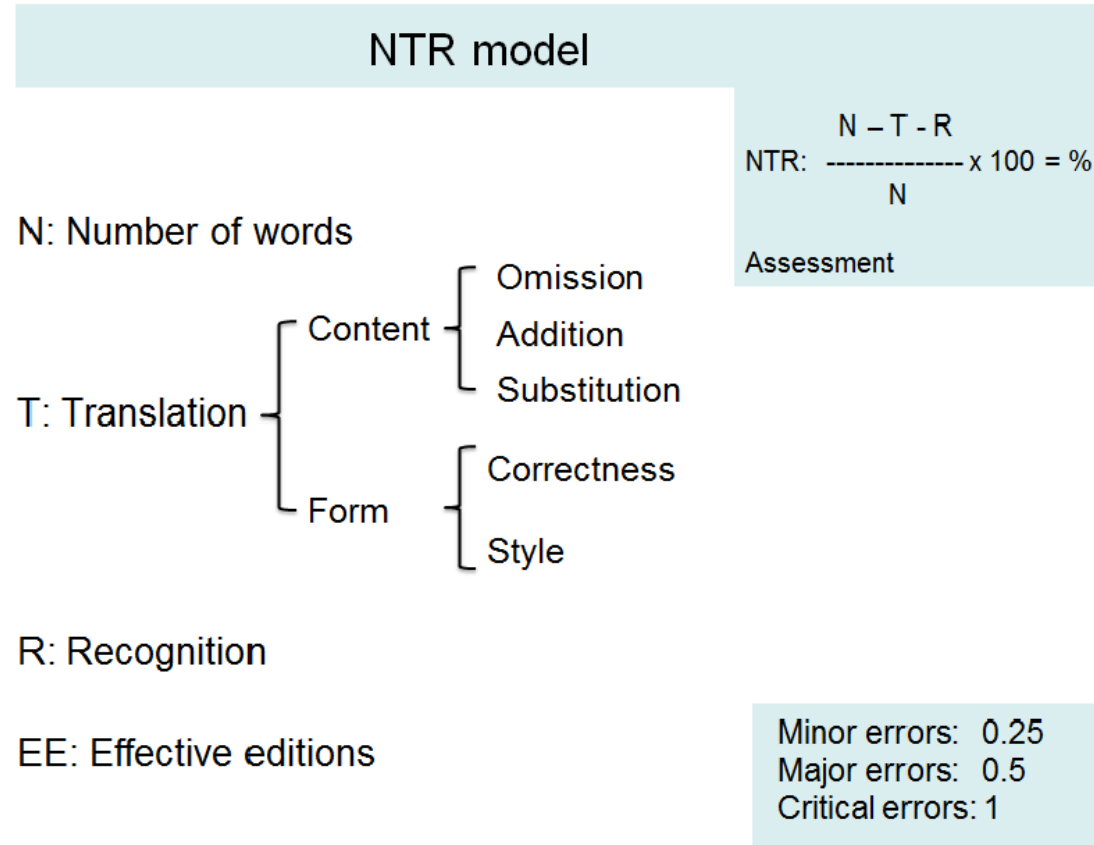


Figure 4: The NTR model (Romero-Fresco & Pöchhacker, 2017)

Example of NTR analysis

Original text (transcribed audio)	Respeaking-based subtitles	Errors
Ladies and gentlemen, I live in Istanbul, which as you probably know is a huge metropolis and in fact one of the biggest cities in the world. As such, it's not a city that is particularly well known for its green spaces and indeed it is a very stressful place to live. However, I count myself very lucky because I happen to have a garden of my own where I can plant flowers and even vegetables, which I then enjoy eating. Now I find that after spending half an hour or an hour or so in the garden I feel much happier and much more relaxed. So any time I'm feeling a bit down, I've got a bit of the blues what I do is I go out and I spend some time digging up some weeds or planting some flowers and this makes me feel much better.	Señores y señoras, vivo en Estambul, una enorme ciudad, de hecho, una de las mayores del mundo. Por lo tanto, no es una ciudad especialmente conocida por sus espacios verdes, de hecho, puede ser un lugar muy estresante para vivir. Sin embargo, me considero muy afortunada, porque tengo un jardín propio, donde puedo plantar flores, e incluso verduras que me encanta comer. Según yo lo veo, después de pasar media hora, o así en el jardín, me siento más contenta y (1) mucho más relajada. Por lo tanto, cada vez que me siento algo deprimida lo que hago es salir y pasar algo de tiempo de Servando (desherbando) (2) o plantando flores. Todo esto me hace sentir mucho mejor.	<p>1.MinT (cont-omiss) (0.25): The viewer would miss some of this description.</p> <p>2. MajR (0.5): The error cannot be recognised.</p> <p>EE: No relevant information is lost here.</p> <p>3. MinR (0.25): The error can be recognised.</p> <p>4. MinT (cont-omiss) (0.25): Omission of a dependent idea unit.</p> <p>5. MinR (0.25): The error can be recognised.</p>

Figure 5: Example of NTR analysis for a test

Example of NTR analysis cont.

Accuracy rate	
MinT: 8 (8 x 0.25 = 2) (cont-omiss) x 6 (cont-add) x 1 (form-corr) x 1 MajT: 0 CritT: 0 Total: 2	MinR: 11 (11 x 0.25 = 2.75) MajR: 1 (1 x 0.5 = 0.5) CritR: 0 Total: 3.25
NTR accuracy rate N = 692 (622 + 70) $\frac{692 - 2 - 3.25}{692} \times 100 = 99.24\% (8/10)$	
EE: 4	
Assessment	
The accuracy rate reaches 98%. The overall quality of the respoken text is excellent. The majority of minor translation errors (8) refer to omissions of dependent idea units (6), which the respeaker could have used as a strategy to keep up with the text. The recognition is acceptable as minor errors (11) are all recognisable and only 1 error causes loss of information. The text is coherent and very easy to follow.	

Figure 5 cont: Example of NTR assessment for a test

Breakdown of professional profiles

46 participants with the following profiles:

- 22% of students had a **clear-cut subtitling**.
- 28% of students had a **clear-cut interpreting**.
- The remaining students had a mixed background of subtitling and interpreting (46%), or no experience whatsoever of subtitling and interpreting (4%).
- Some students (12%) had previous experience of intralingual respeaking.

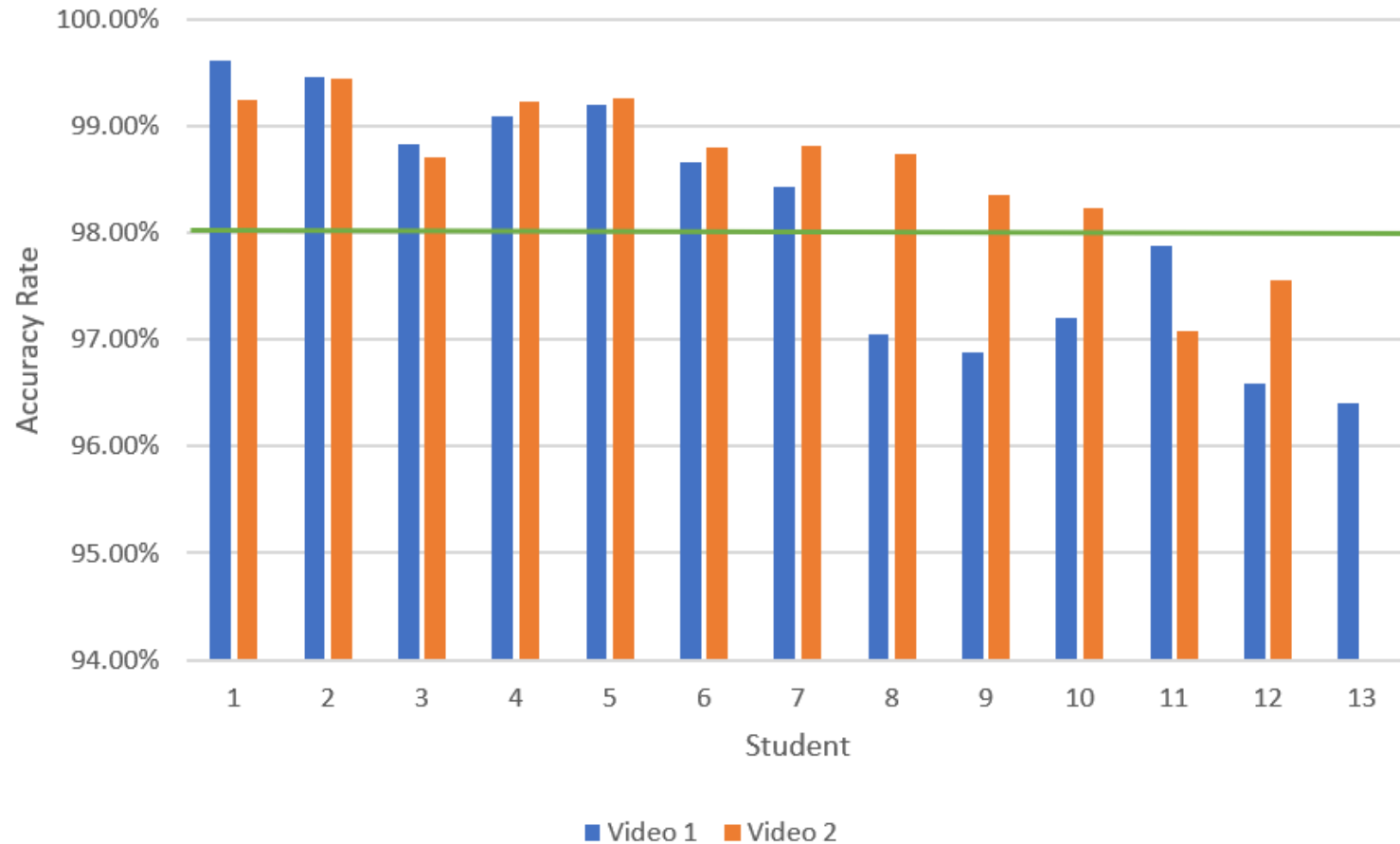
Videos

Week of Course	Intra/interlingual	Title	Genre	Duration	wpm
2	Intralingual	La vida en el arrecife	Documentary	00:05:48	76
2	Intralingual	DELE	Online Class	00:05:16	110
3	Interlingual	Beer	Talking Head	00:05:00	145
3	Interlingual	Médicos sin Fronteras	Interview	00:05:00	125
4	Interlingual	Emma Watson	Speech	00:05:21	107
4	Interlingual	Gardening	Talking Head	00:05:00	159

Interpreters' and subtitlers'

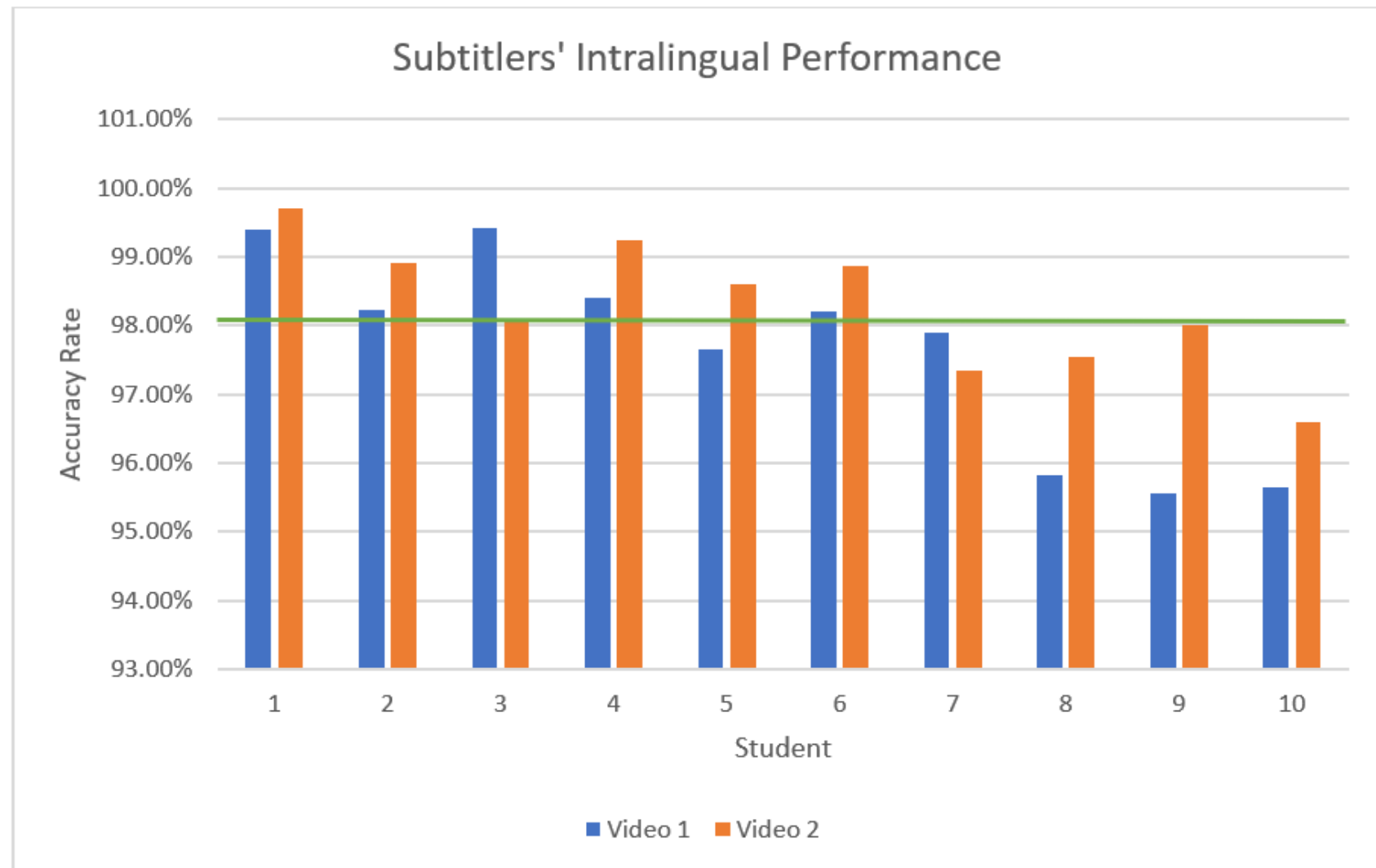
intralingual respeaking performance

Interpreters' Intralingual Performance

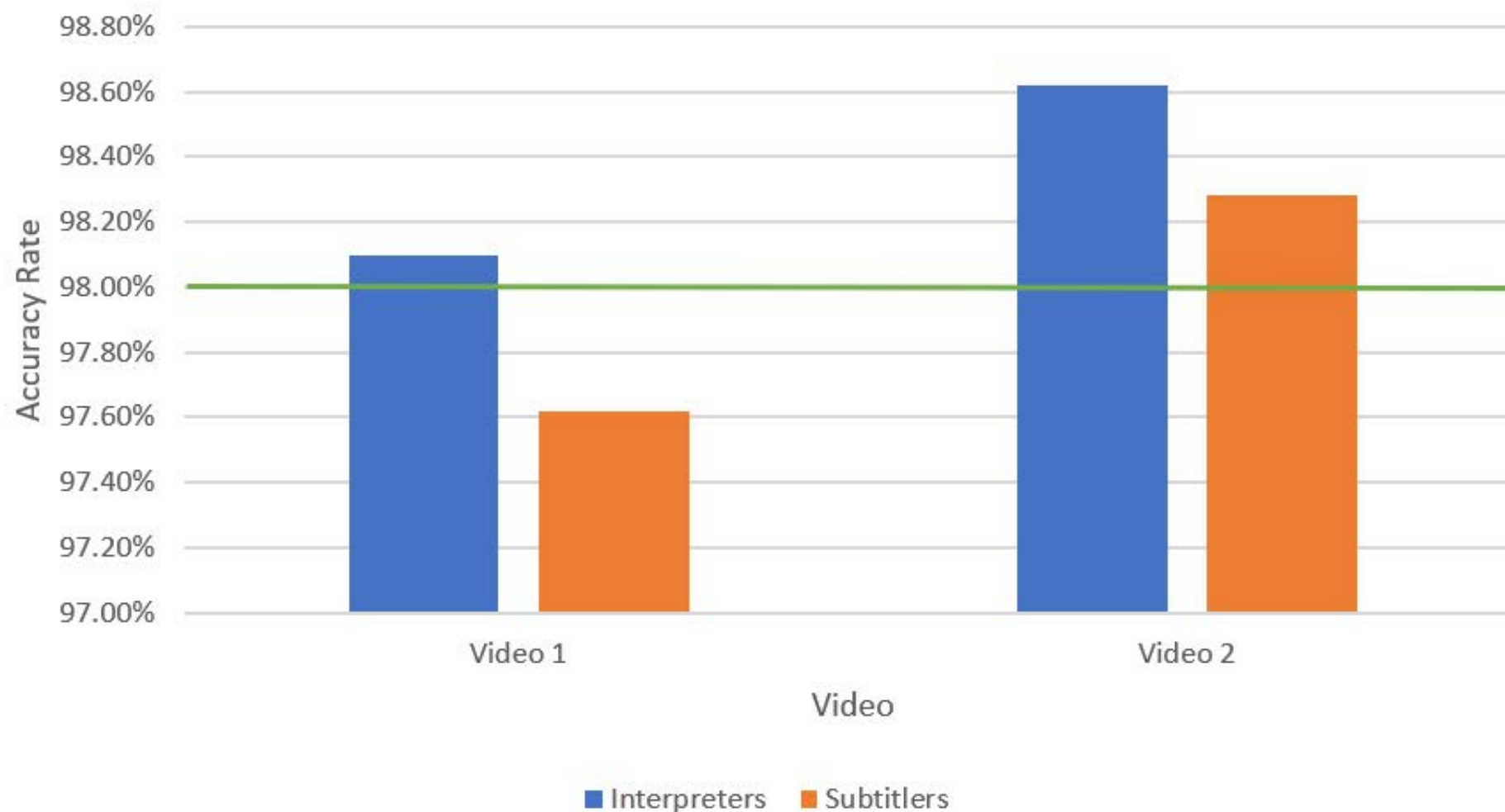


- Average accuracy rate 98.10% in video 1 and 98.62% in video 2 – 98.36% overall.
- 100% of 'good performers' and 16% of 'poor performers' reached 98% in video 1.
- 100% of 'good performers' and 50% of 'poor performers' reached 98% in video 2.
- Edition and recognition errors are balanced.

- Average accuracy rate is 97.62% in video 1 and 98.28% in video 2 – 97.95% overall.
- 50% of subtitlers reached 98% in video 1 and 70% reached 98% in video 2.
- There are some very low accuracy rates of around 95%, which we did not see with the interpreters.



Overall comparison of intralingual performance



Interpreters' and subtitlers'

interlingual respeaking performance

Video 3 – Interlingual ‘Beer – talking head’

Interpreters - Video 3			
Student	T errors	R errors	Accuracy
1	7	7	99.17%
2	7	14	99.04%
3	11	13	98.47%
4	8	17	98.68%
5	14	13	98.80%
6	18	6	98.53%
7	23	29	97.32%
8	11	9	98.19%
9	16	7	98.33%
10	15	19	97.96%
11	19	10	97.08%
12	13	31	96.75%
13	19	17	97.09%
Averages	14	14.7	98.10%

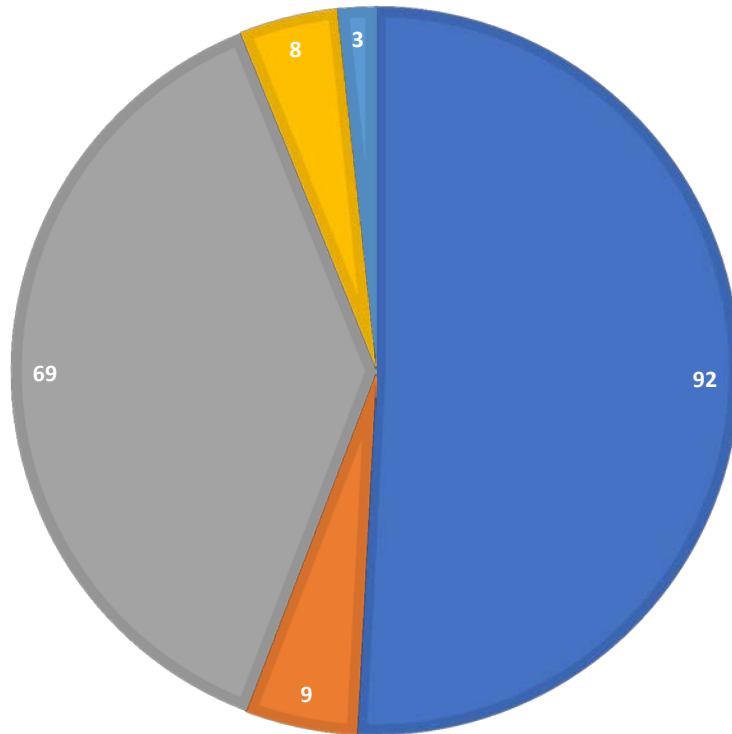
Subtitlers - Video 3			
Student	T errors	R errors	Accuracy
1	10	14	98.65%
2	7	25	98.16%
3	3	30	98.06%
4	22	8	96.74%
5	14	4	98.47%
6	22	20	97.08%
7	11	36	97.05%
8	19	23	95.81%
9	12	22	97.85%
10	9	42	96.71%
Averages	12.9	22.4	97.45%

- 62% of interpreters and 40% of subtitlers reached the threshold of 98%.
- Subtitlers made on average 1.1 fewer T errors, which is a small difference so perhaps not statistically significant.
- Subtitlers' R errors are much higher suggesting they struggle with dictation.

Translation errors

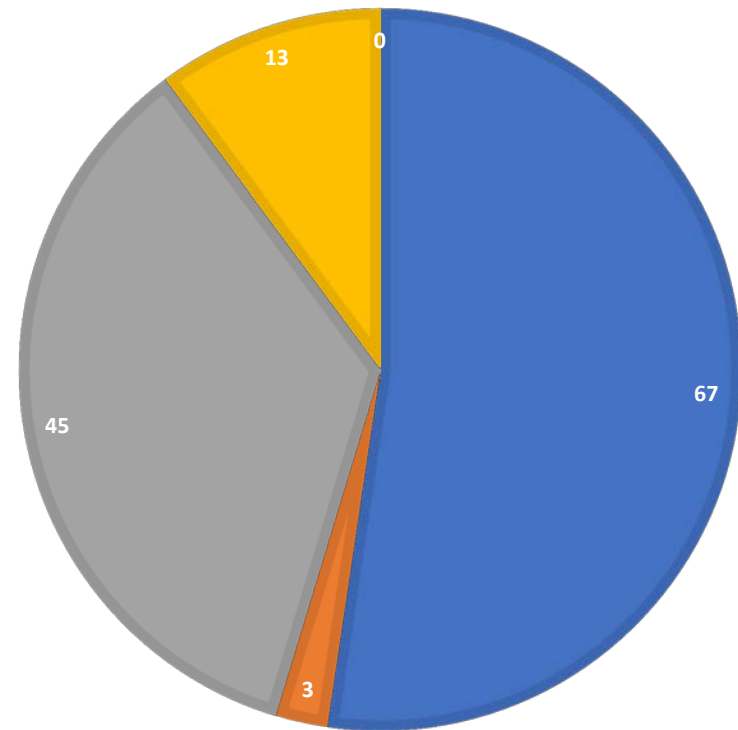
INTERPRETERS VIDEO 3 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



SUBTITLERS VIDEO 3 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



Video 4 – Interlingual ‘MSF interview’

Interpreters - Video 4			
Student	T errors	R errors	Accuracy
1	21	0	98.48%
2	18	6	98.25%
3	8	8	99.19%
4	21	10	97.72%
5	17	10	97.89%
6	18	6	98.06%
7	27	17	97.17%
8	23	2	97.67%
9	22	12	97.28%
10	18	18	97.25%
11	18	5	96.61%
12	31	12	96.08%
13	31	9	93.57%
Averages	21	8.8	97.32%

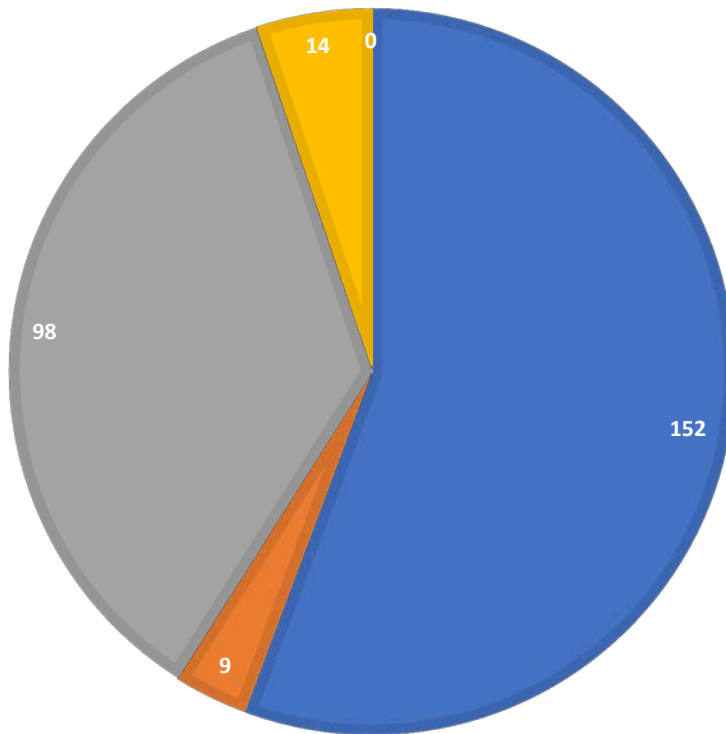
Subtitlers - Video 4			
Student	T errors	R errors	Accuracy
1	13	9	98.51%
2	16	11	97.76%
3	8	17	98.75%
4	28	3	95.95%
5	28	9	96.57%
6	23	4	96.78%
7	23	18	96.90%
8	29	3	97.11%
9	25	11	96.33%
10	23	31	94.58%
Averages	21.6	11.6	96.92%

- 31% of interpreters and 20% of subtitlers reached the 98% threshold.
- Both groups had a high number of translation errors, suggesting students struggled with the specialised terminology.
- Four interpreters and two subtitlers managed to exceed 98% and three others reached at least 97.70%, suggesting that even difficult texts are feasible with little training.

Translation errors

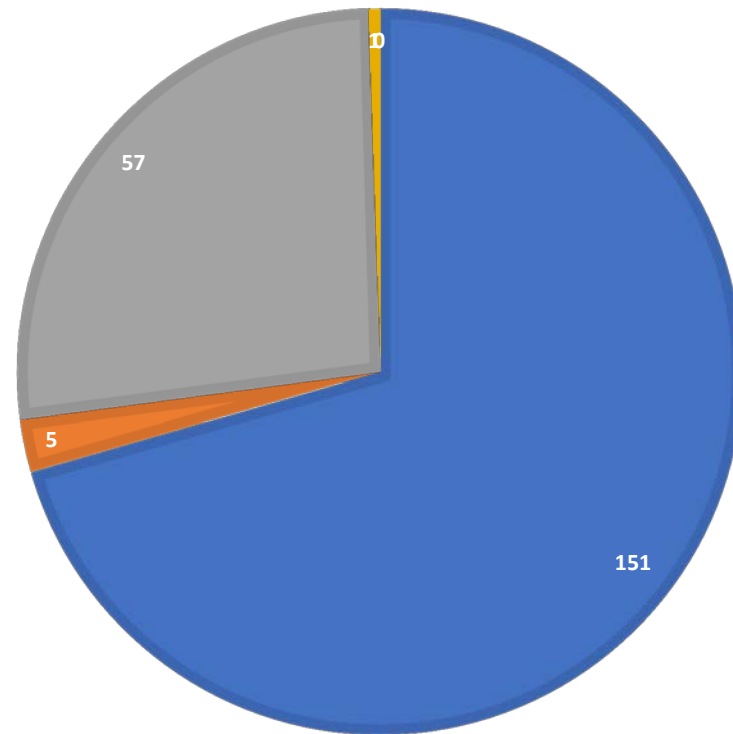
INTERPRETERS VIDEO 4 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



SUBTITLERS VIDEO 4 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



Video 5 – Interlingual Test ‘Emma Watson’

Interpreters - Video 5			
Student	T errors	R errors	Accuracy
1	11	10	98.67%
2	17	12	98.56%
3	10	13	98.64%
4	8	13	98.70%
5	17	15	97.93%
6	18	9	98.65%
7	13	25	97.88%
8	19	13	98.05%
9	24	16	97.47%
10	21	22	95.41%
11	32	23	96.40%
12	24	16	96.20%
13	16	21	96.89%
Averages	17.6	16	97.65%

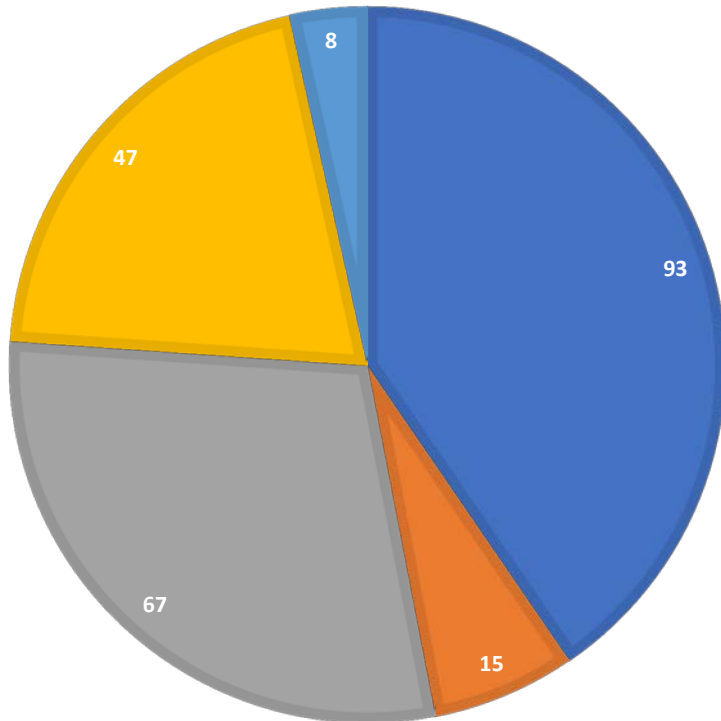
Subtitlers - Video 5			
Student	T errors	R errors	Accuracy
1	9	15	98.92%
2	12	10	98.58%
3	4	32	98.09%
4	26	6	97.34%
5	28	8	96.65%
6	37	16	95.81%
7	12	29	97.46%
8	33	5	96.69%
9	30	13	95.47%
10	17	33	97.16%
Averages	20.8	16.7	97.21%

- 46% of interpreters and 30% of subtitlers reached 98%.
- 5 out of 6 ‘good performing’ interpreters and 1 out of 7 ‘poor performers’ reached the 98% threshold.
- There is a larger difference between good and poor performing interpreters than between interpreters and subtitlers.
- Both groups scored very similar in terms of R errors. Subtitlers had more T errors than interpreters.

Translation errors

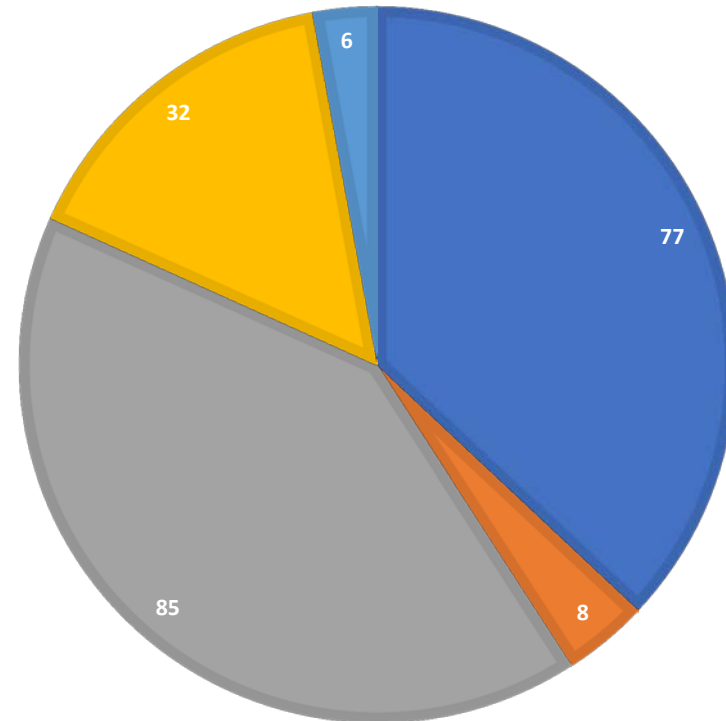
INTERPRETERS VIDEO 5 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



SUBTITLERS VIDEO 5 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



Video 6 – Interlingual Test ‘Gardening’

Interpreters - Video 6			
Student	T errors	R errors	Accuracy
1	11	8	98.81%
2	8	12	99.24%
3	5	15	98.80%
4	13	21	98.58%
5	14	13	98.46%
6	19	7	98.26%
7	14	28	97.66%
8	22	13	97.06%
9	13	14	98.06%
10	11	31	97.18%
12	22	8	97.83%
13	19	21	96.70%
Averages	14.25	16	98.05%

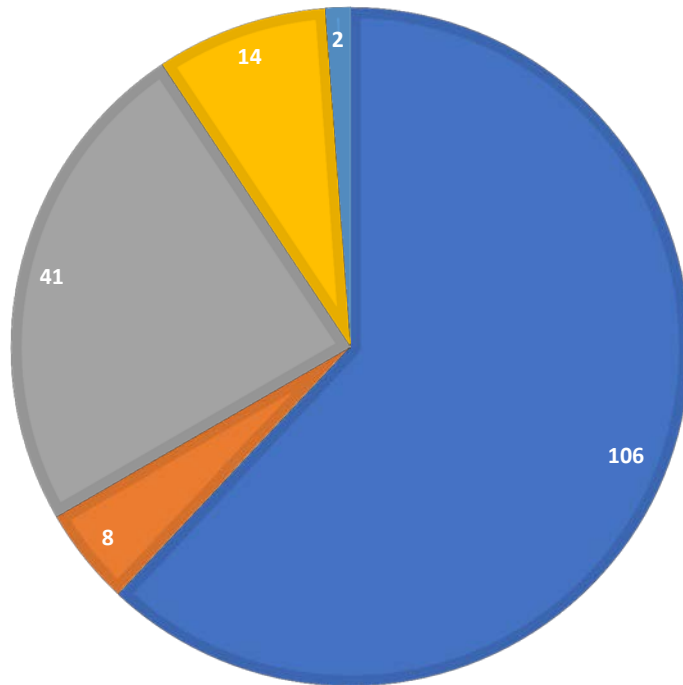
Subtitlers - Video 6			
Student	T errors	R errors	Accuracy
1	11	8	99.16%
2	16	13	98.31%
3	7	44	97.32%
4	15	4	98.46%
5	26	6	97.36%
6	18	15	98.23%
7	12	45	97.28%
8	27	8	96.88%
9	22	25	96.10%
10	13	44	97%
Averages	16.7	21.2	97.61%

- 53% of interpreters and 40% of subtitlers reached 98%.
- All 6 ‘good performing’ interpreters and 1 ‘poor performer’ reached 98%.
- Interpreters found this the second easiest interlingual video and subtitlers found it the easiest video to respeak.
- Some students struggled with recognition still reaching up to 31 errors for interpreters and 45 errors for subtitlers.

Translation errors

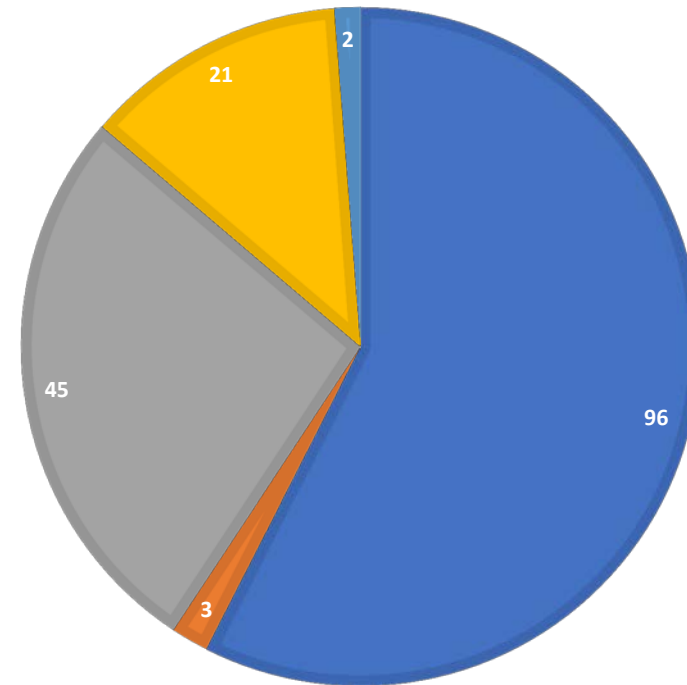
INTERPRETERS VIDEO 5 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style

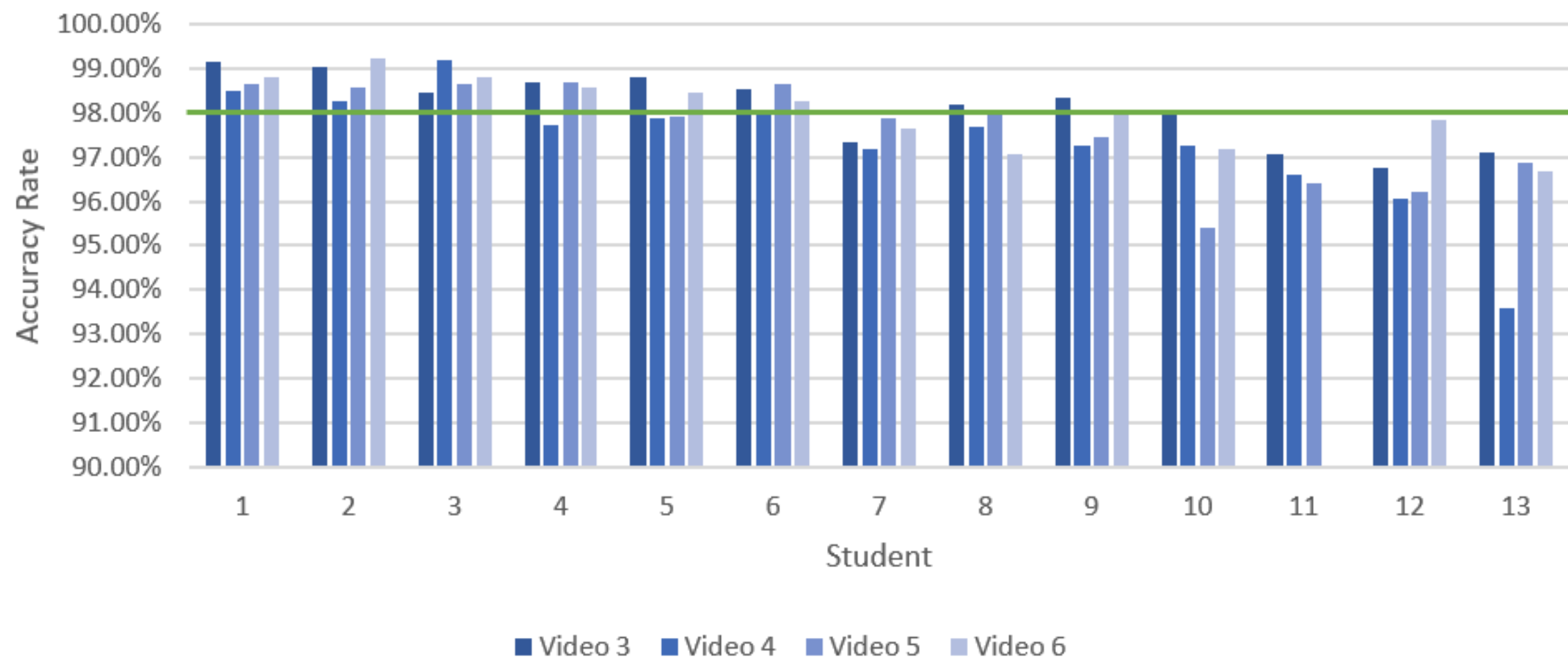


SUBTITLERS VIDEO 5 - T ERRORS

■ Cont-omiss ■ Cont-add ■ Cont-subs ■ Form-corr ■ Form-style



Interpreters' overall interlingual performance



Interpreters' | Overall performance

Video 1	Video 2	Video 3	Video 4	Video 5	Video 6
99.62%	99.25%	99.17%	98.48%	98.67%	98.81%
99.47%	99.45%	99.04%	98.25%	98.56%	99.24%
98.84%	98.71%	98.47%	99.19%	98.64%	98.80%
99.09%	99.24%	98.68%	97.72%	98.70%	98.58%
99.20%	99.27%	98.80%	97.89%	97.93%	98.46%
98.67%	98.80%	98.53%	98.06%	98.65%	98.26%
98.44%	98.82%	97.32%	97.17%	97.88%	97.66%
97.05%	98.75%	98.19%	97.67%	98.05%	97.06%
96.88%	98.36%	98.33%	97.28%	97.47%	98.06%
97.21%	98.23%	97.96%	97.25%	95.41%	97.18%
97.88%	97.08%	97.08%	96.61%	96.40%	N/A
96.59%	97.56%	96.75%	96.08%	96.20%	97.83%
96.41%	N/A	97.09%	93.57%	96.89%	96.70%
Averages					
98.10%	98.62%	98.10%	97.32%	97.65%	98.05%

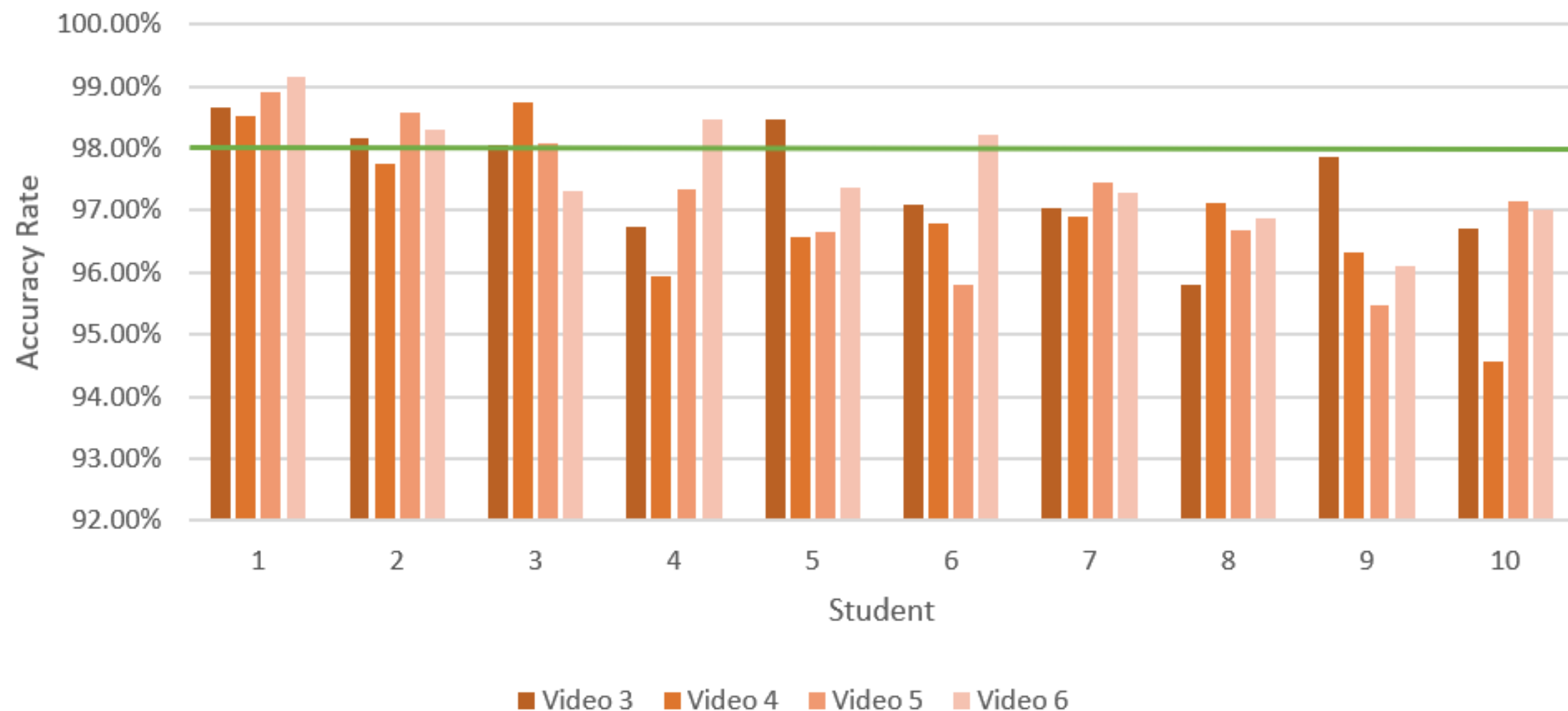
The interpreters produced 76 respoken texts of which the following met or exceeded the 98% threshold:

- 17/25 (68%) intralingual texts
- 25/51 (49%) interlingual texts
- 13/25 (52%) interlingual tests

Video 3 had fewer average translation errors at 14 per text.

Video 4 had fewer recognition errors, with 8.8 errors per text. This was the most difficult video to translate live due to specialised terminology. Students may have decided to focus on dictation to control their errors.

Subtitlers' overall interlingual performance



Subtitlers' | Overall performance

Video 1	Video 2	Video 3	Video 4	Video 5	Video 6
99.40%	99.71%	98.65%	98.51%	98.92%	99.16%
98.22%	98.90%	98.16%	97.76%	98.58%	98.31%
99.41%	98.06%	98.06%	98.75%	98.09%	97.32%
98.41%	99.25%	96.74%	95.95%	97.34%	98.46%
97.64%	98.59%	98.47%	96.57%	96.65%	97.36%
98.21%	98.86%	97.08%	96.78%	95.81%	98.23%
97.89%	97.34%	97.05%	96.90%	97.46%	97.28%
95.81%	97.55%	95.81%	97.11%	96.69%	96.88%
95.56%	98%	97.85%	96.33%	95.47%	96.10%
95.65%	96.60%	96.71%	94.58%	97.16%	97%
Averages					
97.62%	98.28%	97.45%	96.92%	97.21%	97.61%

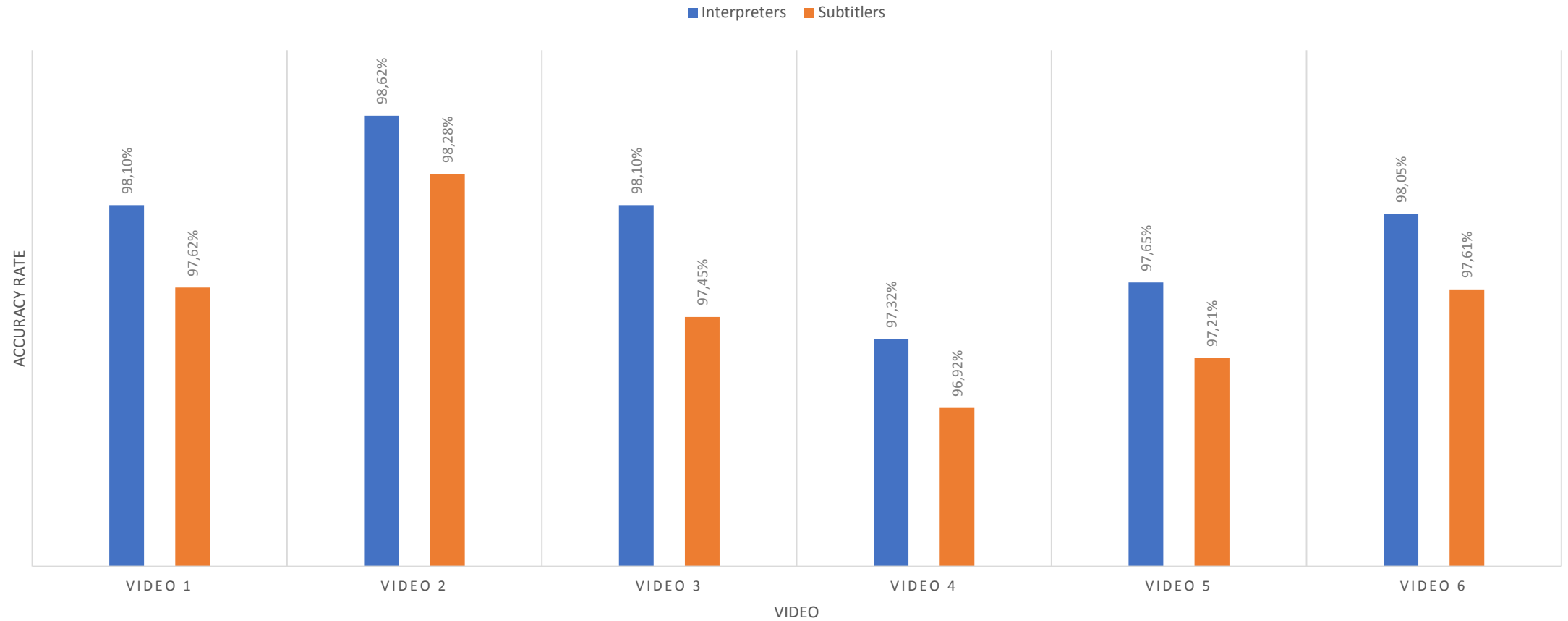
The subtitlers produced 60 respoken texts of which the following met or exceeded the 98% threshold:

- 12/20 (60%) intralingual texts
- 13/40 (32%) interlingual texts
- 7/20 (35%) interlingual tests

This suggests subtitlers cannot cope with the element of live translation, which is required for interlingual respeaking.

Some subtitlers can be good interlingual respeakers, but perhaps when they are taken as a group there is no guarantee they can be good respeakers.

Overall accuracy rates of interpreters and subtitlers



Conclusions

- For the interlingual tests, 50% of interpreters reached the 98% threshold, while only 1 in 5 subtitlers reached 98%.
- Interpreters make consistently fewer R errors than subtitlers with an average of 4.1 fewer errors per text.
- Differences in T errors are much closer for both groups, with interpreters making on average 1.8 fewer errors than subtitlers.
- Form errors do not seem to pose any problems, content errors do – especially omissions and substitutions:
 - Interpreters had an average of 8.5 omissions per text, subtitlers had 9.8 omissions per text.
 - Interpreters has an average of 5.2 content substitutions per text, subtitlers had an average of 5.8.

Conclusions cont.

- In terms of error severity, both groups maintained a similar pattern of making more minor, then major then critical errors.
- There are a few differences: subtitlers made more errors in general; interpreters made more critical content substitutions than major substitutions and the poor performing interpreters had a slightly greater difference between major and critical substitution errors with a difference of 0.9, good interpreters had an average difference of 0.7.

Conclusions cont.

- ILS seems feasible (97.6%, 4/10)
- Interpreters perform better than subtitlers
- There is a greater difference between 'good' and 'poor' performing interpreters than there is between interpreters and subtitlers.
- Interpreter \neq good performer / Subtitler \neq poor performer
- Translation and recognition are equally important and challenging
- Good performers have around 50% fewer translation and recognition errors than bad performers, including consistently less serious errors.
- Bad performers struggle to keep up and as a result omit too many full sentences, mistranslate the source text and dictate less clearly.
- Subtitlers seem to struggle trying to keep up with the text, as a result they have more omissions, more mistranslations and more recognition errors.

ILSA Interlingual Live
Subtitling for Access



UniversidadeVigo



Thank you for your attention