

## Apples and oranges?

# A Comparative Study of Accuracy in Live Subtitling: The NER Model

# Accuracy in Live Subtitling

- Quality in live subtitling
  - Delay
  - Accuracy

# Accuracy in Live Subtitling

97-98% accuracy



# Basic requirements for a model

- 1) Functional and easy to apply
- 2) Include the basic principles of WER calculations in SR
- 3) Different programmes, different editing
- 4) Possibility of edited and yet accurate respeaking
- 5) Compare subtitles with original spoken text
- 6) Include other relevant info (delay, position, speed)
- 7) Provide both percentage and food for thought in training

# Traditional WER methods

- US National Institute of Standards and Technology

- Accuracy Rate  $\frac{N - \text{Errors}}{N} \times 100 = \%$

Where	is	the	whole	wheat	flour	
		<b>D</b>	<b>S</b>	<b>S</b>	<b>I</b>	<b>S</b>
Where	is		hole	we	eat	flower

# Traditional WER methods

- US National Institute of Standards and Technology

- Accuracy Rate  $\frac{N - D - S - I}{N} \times 100 = \%$

# Traditional WER methods

- US National Institute of Standards and Technology

- Accuracy Rate  $\frac{6 - 1 - 3 - 1}{6} \times 100 = 16\%$

- But...

Well, you know, you have to try and put out a good performance, I mean, yeah, it's kind of a stepping stone, isn't it, really?

You have to try to put out a good performance. It's a stepping stone.

# Basic requirements for a model fulfilled so far

## **Yes**

Functional and easy to apply

Include the basic principles of WER calculations in SR

Compare subtitles with original spoken text

## **No**

Possibility of edited and yet accurate respeaking

Different programmes, different editing

Include other relevant info (delay, position, speed)

Provide both percentage and food for thought in training



# The US



## Caption Accuracy Metrics Project

Caption Viewer Survey:  
Error Ranking of Real-time Captions  
in Live Television News Programs



Automatic Measure of Caption Quality

(steno – verbatim approach)

- WGBH: “There is a wide range of error types in real time captioning and they are not all equal in their impact to caption viewers” .

“Treating all errors the same does not provide a true picture of caption accuracy” .

# NER MODEL

$$\text{Accuracy} = \frac{N - E - R}{N} \times 100 = \%$$

**Correct editions:**

**Assessment:**

**Target = 98%**

# NER Model –lessons to be learnt

- Types of errors (feedback from DTV4ALL project)
  - 1) “There are errors, yes, but you can easily figure out what the correct form was meant to be.  
I’ m deaf, not stupid”.
  - 2) “Thanks to live subtitling I’ m now bilingual.  
I can speak English and teletext”.
  - 3) “As far as I’ m concerned, they are not errors, but lies”

# NER Model – shortcomings – lessons to be learnt

- Types of errors (feedback from DTV4ALL project)
  - 1) Minor edition or recognition errors (0.25)  
(meaning still clear)
  - 2) Standard edition or recognition errors (0.5)  
(loss of information)
  - 3) Serious edition or recognition errors (1)  
(misinformation/lie)

## Minor Recognition Errors (meaning still clear)

**What a great goal by a Ryan Giggs!**

**Simon brown has been appointed  
new chairman of Rolls Royce.**

**For people are still missing  
following Sunday's tornado.**

## Minor Recognition Errors (meaning still clear)

What a great goal **by Ryan Giggs!**

Simon **Brown** has been appointed  
new chairman of Rolls Royce.

**Four** people are still missing  
following Sunday's tornado.

# Standard Recognition Errors (nonsense)

**He's a buy you a bull asset.**

**Is it really attend Tatian?**



# Standard Recognition Errors (nonsense)

He's a buy you a bull asset.

He's a **valuable** asset.

Is it really attend Tatian?

Is it really **a temptation**?

# Serious Recognition Errors (misinformation/lie)

**Public funding for universities  
has been cut by 15% this year.**

**He never talks dirty.**

# Serious Recognition Errors (misinformation/lie)

**Public funding for universities  
has been cut by 15% this year.**

**Public funding for universities  
has been cut by 15% this year.**

**He never talks dirty.**

**He never talks to Rudy.**

# Edition errors

## Independent and dependent idea units

Independent idea unit: often a sentence, it may be composed of several dependent idea units and it makes sense as a full message.

Dependent idea unit: often a complement, it provides information about the “when”, the “where”, the “how”, etc. of an independent idea unit.

# Edition errors

## Independent and dependent idea units

The two victims of a fatal fire in Melbourne are yet to be formally identified.

The blaze started this morning at the front of the house.

# Edition errors

## Independent and dependent idea units

Independent IU



The two victims of a fatal fire in Melbourne are yet to be formally identified.

The blaze started this morning at the front of the house.

# Edition errors

## Independent and dependent idea units

Independent IU

The two victims of a fatal fire in Melbourne are yet to be formally identified.

The blaze started this morning at the front of the house.

Independent IU

# Edition errors

## Independent and dependent idea units

Independent IU

The two victims of a fatal fire in Melbourne are yet to be formally identified.

The blaze started this morning at the front of the house.

Dependent IU

Dependent IU

Independent IU



# Edition errors in the NER model

Serious error (1): edition that causes change of meaning (i.e. misinformation, lie).

Standard error (0.5): omission of a full independent IU or omission of a dependent IU that renders the whole unit meaningless/nonsensical.

Minor error (0.25): omission of a dependent IU that doesn't render the whole unit meaningless/nonsensical.

# NER MODEL

$$\text{Accuracy} = \frac{N - E - R}{N} \times 100 = \%$$

**Correct editions:**

**Assessment:**

**Target = 98%**

# Current use of the NER model

- Included in the official UNE subtitling guidelines in Spain
- Used by Ofcom to assess the quality of live subtitles in the UK
- Used by the BBC, ITV, Channel 4, Channel 5, Sky in the UK
- Adopted by the Brazilian regulator ABNT in its updated official subtitling guidelines NBR15290
- Under discussion for inclusion by the official Canadian regulator CRTC in the new official guidelines

# Current use of the NER model

- Used by the leading subtitling companies in the world for training and assessment purposes (Ericsson, Ai-Media, Deluxe, Verbavoice)
- Used by the leading broadcaster and subtitling company in Switzerland (SwissTXT)
- Used by regulators, broadcasters, subtitling companies and universities in other countries such as France, Italy, Belgium, Holland, Poland, Germany, Austria or South Africa.

# Current use of the NER model

“We have chosen the NER model because it is anchored in robust independent consumer research and produces quantitative scores that are consistent with viewers’ quality perceptions.”

(Robert Scott, Ai-Media)

“The NER model is the best system for measuring quality in live captioning”

(David Keeble, CRTC, Canada)

## Current use of the NER model

“The NER system is probably the best way yet devised to measure the accuracy of live captioning, in that it is repeatable and, therefore, usable across various suppliers and broadcasters as well as being adopted in overseas territories. Repeatable, that means that we are all kicking towards the same goalposts, we are all singing from the same hymnbook. It's an objective assessment; it's not a subjective assessment, which can be swayed by other reviewers' points of view. So, as a repetitive of a broadcaster, I think the NER model is great”.

(Paul Richardson, Media Services Manager at Seven Network)

# Orange and apples



NER

Orange with apples





## Apples and oranges?

# A Comparative Study of Accuracy in Live Subtitling: The NER Model