8th Live Subtitling and Accessibility Symposium

The quality of automatic and human live captions in English – and beyond

Pablo Romero-Fresco and Nazaret Fresno
(GALMA, Universidade de Vigo / The University of Texas at Rio Grande Valley)

UNIVERSITY OF LEEDS

Universidade Vigo

SWISS TXT

INTERSTENO

TRANS MEDIA CATALONIA

Gion’s questions

- Why is the NER model so time-consuming?
- Is it really scalable?
- Won't AI soon sweep away everything we're talking about here?

QuaLiSub

The Quality of Live Subtitling

1 - To assess the accuracy of ASR in Galician, Basque and Catalan for live subtitling.

2 - To assess the accuracy of human and automatic live subtitles in English and Spanish.

3 - To explore the automation of the NER model

4 - To monitor the quality of the automatic ASR and MT tool used by the EU Parliament

Universidade Vigo

UNIVERSITAT JAUME I

Universidad del País Vasco · Euskal Herriko Unibertsitatea

The University of Texas Rio Grande Valley

rtve

QuaLiSub
The Quality of Live Subtitling

Language	Minutes analysed	Model
Galician	475	WER and NER
Basque	225	NER
Catalan	410	WER and NER
Spanish	1000	WER and NER
English	800	WER and NER
Total	**2910** **(= 61000 subtitles)**	

The quality of human and automatic subs in English

- 2018-2023
- 17.000 captions
- 800 minutes of live captions
  - 388m human (steno and respeaking)
  - 410m ASR
- Sky (UK), VITAC and ENCO (UK), CRTC (Canada)
- NER model

The quality of human and automatic subs in English

Case Study	Human average	Automatic average	Year
Case Study 1 (Sky)	97% (2.5/10)	95.7% (0/10)	2018
Case Study 2 (Vitac)	98.8% (7/10)	96.3% (1/10)	2020
Case Study 3 (Enco)	99.4% (8.5/10)	96.3% (1.5/10)	2021
Case Study 4 (Canada)	98.8% (7/10)	97.8% (4.5/10)	2020-2022
Totals	**98.5% (6.5/10)**	**96.5% (1/10)**	

The quality of human and automatic subs in English

Jim Pattison, 2021-2022: 98.3% (31 samples)

Miracle Channel, 2021-2022: 98.3% (24 samples)

Case Study 4 (Canada)	98.8% (7/10)	97.8% (4.5/10)	2020-2022

The quality of human and automatic subs in English

Automatic captions:

- Close to 98%
- Slightly over 98% with human assistance
- Paves the way for other languages
- Still issues with punctuation, proper nouns, numbers, speaker ID
- Always (near) verbatim, so often fast and difficult to read

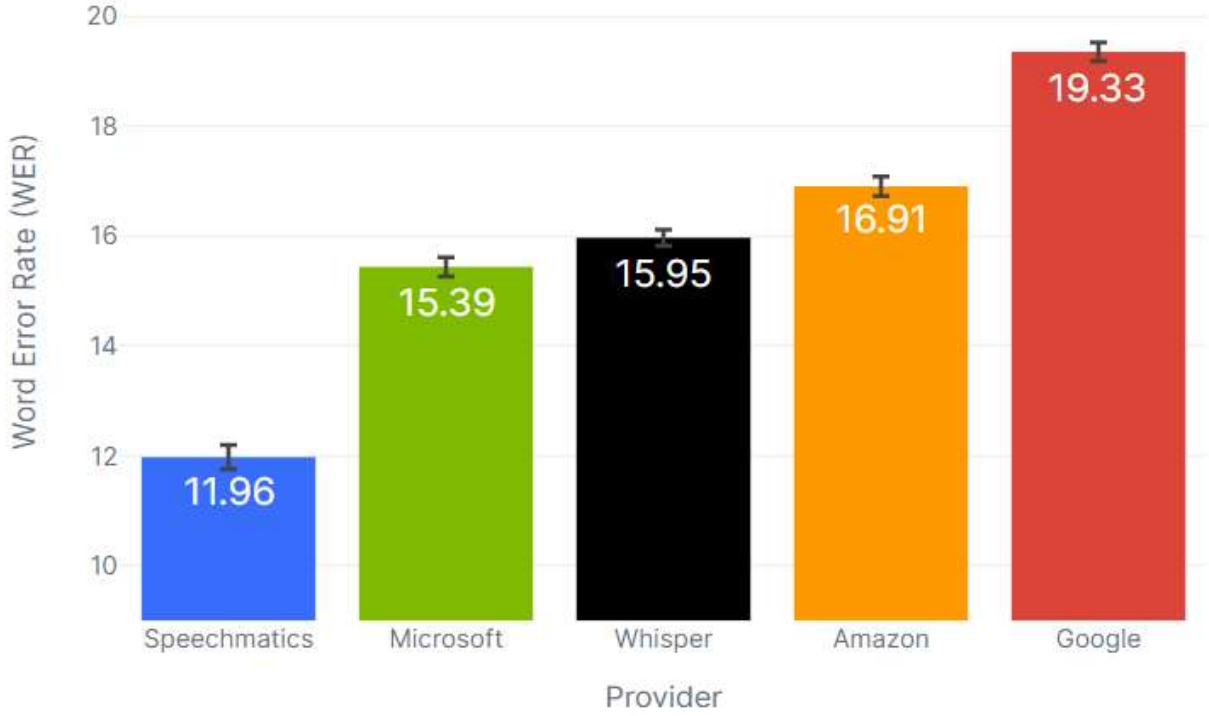
The quality of human and automatic subs in English

Human captions:

- Excellent (steno) or very good (respeaking)
- More accessible: fewer errors and more readable speeds

The quality of human and automatic subs in English

The WER War


20
18
16
14
12
10
Word Error Rate (WER)
11.96
15.39
15.95
16.91
19.33
Speechmatics
Microsoft
Whisper
Amazon
Google
Provider

The WER War

** We are aware of the limitations of WER, one major issue being that errors involving misinformation are given the same weight as simple spelling mistakes. To address this, we normalize our transcriptions to reduce penalties for differences in contractions or spelling between British and American English that humans would still consider correct. Going forward, we intend to adopt a metric based on NER.

The WER War

Blog - Technical

Oct 14, 2022 | Read time 4 min

The Problem with Word Error Rate (WER)

In part one of a two-part piece, Accuracy Team Lead, John Hughes, explains how and why Word Error Rate as an accuracy measure is outdated and often misaligned with human judgment. Part Two looks at where we go next.

The Problem with Word Error Rate (WER)

John Hughes, Accuracy Team Lead

John Hughes
Accuracy Team Lead

WER vs NER

	Speechmatics		Whisper		Pre-AI ASR	
English speech						
English interview						
Spanish speech						
Spanish interview						

WER vs NER

	Speechmatics		Whisper		Pre-AI ASR	
English speech	3.6%		3.8%		3%	
English interview	15.7%		19.1%		25.7%	
Spanish speech	1.7%		2%		5.2%	
Spanish interview	14%		10.3%		27.6%	

WER vs NER

wasn t that your lucky day that he was nearby absolutely ok(okay) so first of all mark tell me what happened i was responding to a person with a weapon call my lights and sirens going(go) came(on) up(came) to(into) an intersection was involved in an auto accident the next thing i know is my vehicle s(is) close to a sidewalk my airbag s(is) blown and there s(was) a fire right in front of me i had a so you don t remember impact i don t i don t remember impact i just kind of woke up basically still i knew something was wrong i couldn(could) t(not) move around too much joe came over started(said) yo hey your car s(is) on fire yeah and i was kind of out of it like(kinda) well thanks he told the radio this is where i m(am) at and i need help so you didn t know the car was on fire i did i did see that the car was on fire but it didn(did) t(not) register enough to know that i was in danger that you should get out exactly yeah that s my natural instinct when the car s on fire it s like not just to watch it but clearly you were like out of it so what happened so where were you and how did you know this happened i was actually visiting my sister in south philly and at the time we were watching tv just hanging out and we heard this loud boom and i was just like well(wow) what the heck was that so we all got up ran to the door looked out and it was kind of smoky but right away i knew it was a cop car right i m(am) a volunteer firefighter for 3(three) and 1(a) 2(half) years look at that so when i realized(realised) it was a cop it kind of hit home a little bit for me so instantly i just ran up(off) to the car i ran up to him i was like how are you and he was(is) all disoriented he didn t know where he was and i said your car s(is) on fire and he s like just staring(starring) at me i m(and) just(was) like i m going to have to get you out(yeah) and i tried to open the door and the door wasn t opening so i m like i m going to have to(gonna) get you out through(of) the window and he just threw his arms out and i just dragged him out but that s like just dead weight because he wasn t really moving so you had to lift him out and over that door right correct yeah he was heavy wow yeah it s all muscle it s all solid muscle so yeah that s really impressive that you could do that and obviously your training as a firefighter comes in handy right there too yeah somewhat it does kind of but more or less the instinct to go and help somebody was from my dad he passed away two months ago and he gave me he inspired me to do great things his famous saying to me was if someone needs help help them doesn t matter what the circumstances are if you have 10 dollars and a guy on the street who needs help you(needed) need(10) 10(dollars) give it to him and that s something exactly what he would do so he inspired me he sounds like a great man i m sorry for your loss and you re a good guy joe first of all what happened what s wrong with you the leg is you know it s still banged up i still have tests and doctors i have to go see about it i m just lucky that joe was there otherwise i d still be sitting in that car because i think after you got him out the whole car caught fire right within a minute yeah crazy just like in that(the) video i mean that video was i ll(d) say two minutes after when the fire department(firefighter) got there but it was burning for a good time so he saved your life absolutely this guy absolutely there s really nothing left of the car yeah and(so) you re pals now right absolutely what are you going to do after you graduate i m more or less thinking about the military service right out of(at) high school i don t know which branch yet but after the military service i was thinking about a firefighter or a police officer we can t legally give you a gift we can t do that because you re an officer of the law i don(do) t(not) understand that law but you it(know) seems like when you do something and you want to help somebody your girlfriend s in the audience and she must be exhausted from taking care of you all the time yep right all right so because i think she s exhausted from taking care of you she deserves a vacation so she s going to get a six night stay for two at the cebu(ceblue) villas and beach resort in anguilla she ll stay in a three bedroom villa with its own pool she s going to(gonna) have a picnic on a private island and she s going to(gonna) get a couples massage so i don(do) t(not) know who she s going to(gonna) bring but that s where she s going she s going to bring joe you re a great buddy all right you could(can) bring joe joe we heard that you ride your bicycle to work a lot but your bike has a flat so it always has a flat tire(tyre) so we got you a new bike right here and yes yes yes that s awesome and we know that when you aren(re) t(hard) working as a volunteer firefighter you like to play video games so we re going to give you this brand new entertainment system over here oh my god oh my god oh my god oh my god oh my god oh my god oh my god oh my god oh my(sh) god(t) congratulations thank you you re welcome that s awesome thank you

177 errors

WER vs NER

saying to me was, if someone needs help, help them. Doesn't matter what the circumstances are. If you have $10 and a guy on the street who needs help, you need $10, give it to him. And that's

something exactly what he would do. So he inspired me. He sounds like a great man. I'm sorry for your loss. And you're a good guy, Joe. First of all, what's wrong with you? The leg is still banged up. I still have tests and doctors I have to go see about it. I'm just lucky that Joe was there. Otherwise, I'd still be sitting in that car. Because I think after you got him out, the whole car caught fire, right? Within a minute, yeah. Crazy. Just like in that video. I mean, that video was, I'll say, two minutes after when the fire department got there. But it was burning for a good time. So he saved your life. Absolutely. This guy. There's really nothing left of the car. Yeah. And you're pals now, right? Absolutely. What are you going to do after you graduate? I'm more or less thinking about the military service. Right out of high school. I don't know which branch yet. But after the military service, I was thinking about a firefighter or a police officer. We can't legally give you a gift. We can't do that. Because you're an officer of the law. I don't understand that law. But it seems like when you do something and you want to help somebody, your girlfriend's in the audience. And she must be exhausted from taking care of you all the time. Yep. Right? All right, so because I think she's exhausted from taking care of you, she deserves a vacation. So she's going to get a six-night stay for two at the Cebu Villas and Beach Resort in Anguilla. She'll stay in a three-bedroom villa with its own pool. She's going to have a picnic on a private island. She's going to get a couples massage. So I don't know who she's going to bring, but that's where she's going. She's going to bring Joe. You're a great buddy. All right. You could bring Joe. Joe, we heard that you ride your bicycle to work a lot, but your bike has a flat. So it always has a flat tire. So we got you a new bike right here. And. Yes. Yes. Yes. That's awesome. And we know that when you aren't working as a volunteer firefighter, you like to play video games. So we're going to give you this brand new entertainment system over here. Oh, my God. Oh, my God. Oh, my God. Oh, my God. Oh, my God. Oh, my God. Oh, my God. Oh, my God. Congratulations. Thank you. You're welcome. That's awesome. Thank you.

2 errors

WER vs NER

	Speechmatics		Whisper		Pre-AI ASR	
English speech		99.6%		99.7%		98.9%
English interview		99.4%		99.9%		97%
Spanish speech		99.2%		99.9%		97.1%
Spanish interview		99.1%		99.7%		95%

WER vs NER

Can NER be automated?

- ❑ 300 examples of NER
- ❑ Distinction between minor, standard and serious errors
- ❑ 3 months
- ❑ 2 iterations

Can NER be automated?

Confusion matrix:

GPT Score	Human Label					
	0.25	0.5	0.75	1	1.25	All
0.25	79	26	0	6	1	112
0.5	39	24	3	4	0	70
0.75	11	3	1	0	0	15
1	33	41	4	14	0	92
1.25	4	3	2	1	0	10
1.5	5	1	0	0	0	6
1.75	1	2	0	0	0	3
2	3	2	1	1	0	7
2.25	1	0	0	0	0	1
All	176	102	11	26	1	316

Can NER be automated?

- 40% success rate
- Good with minor errors
- Sometimes struggles to differentiate standard errors (meaning is lost) from serious errors (meaning is changed)

Conclusions (2017-2022)

- 2017- late 2022: improvement of autosubs from unusable to useable
- Human captions still better and more accessible (edited)

Conclusions (2023)

- New AI-powered ASR: potentially more accurate than human*

* verbatim, no or worse speaker ID

- Important decisions adopted on the basis of unreliable WERs
- Potential automation of NER
- Future of intra and interlingual human and automatic subs?

Gion’s questions

- Why is the NER model so time-consuming?
- Is it really scalable?
- Won't AI soon sweep away everything we're talking about here?

8th Live Subtitling and Accessibility Symposium

The quality of automatic and human live captions in English – and beyond

Pablo Romero-Fresco and Nazaret Fresno
(GALMA, Universidade de Vigo /
The University of Texas at Rio Grande Valley)

UNIVERSITY OF LEEDS

Universidade Vigo

SWISS TXT

INTERSTENO

TRANS MEDIA CATALONIA