

PROBLEMATICA ȘI LIMITELE SUBTITRĂRII AUTOMATE

Lect. univ. dr. Cristina VARGA

Universitatea Babeș-Bolyai, Cluj-Napoca

10-13 aprilie 2023, Timișoara, Conferința SmartDiaspora

OBIECTIVELE CERCETĂRII

Cunoașterea:

- ▶ structurii programelor de subtitrare automată;
- ▶ varietății și tipologiei acestora;
- ▶ problemelor cu care utilizatorul se confruntă atunci când le utilizează;
- ▶ situațiilor în care acestea pot fi folosite cu succes;
- ▶ situațiilor în care folosirea acestora nu este recomandabilă;
- ▶ evoluției TA în domeniul TAV.

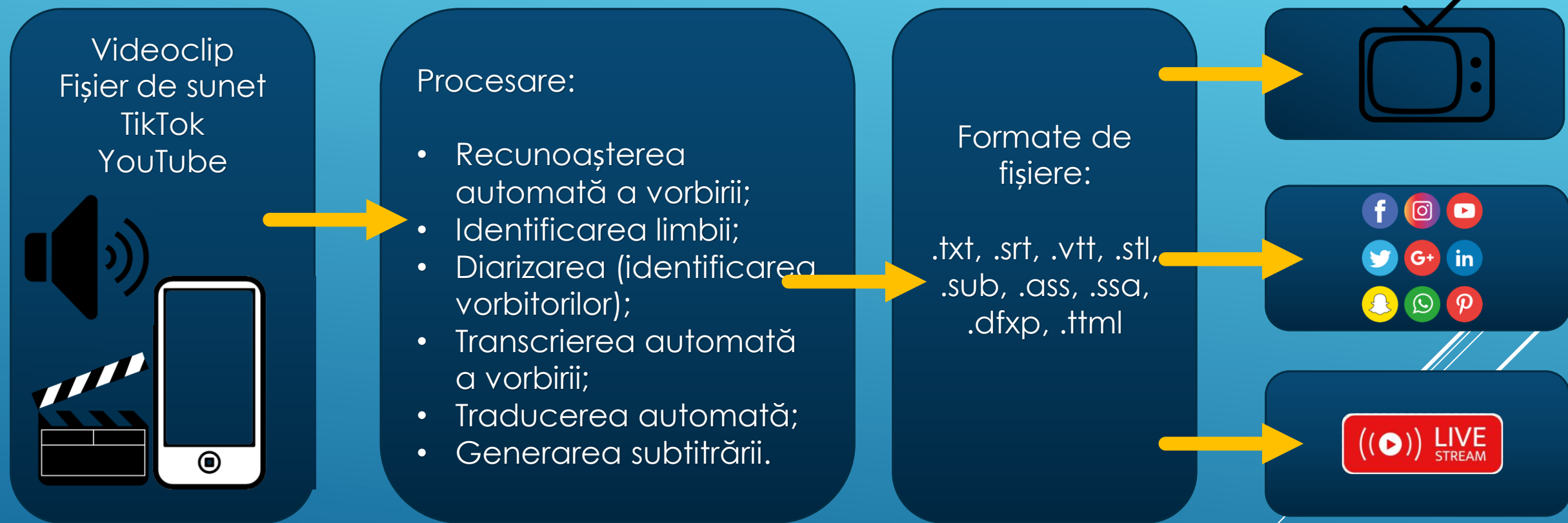
METODOLOGIA DE LUCRU

1. Fluxul de lucru al programului de subtitrare automată:
transcriere > traducere > subtitrare
2. Analiza fiecărei etape pentru a putea observa:
 - ▶ care sunt punctele slabe ale fiecărui program;
 - ▶ care sunt limbile în care calitatea subtitrării este mai slabă;
 - ▶ ce categorii de erori sunt mai frecvente în transcriere;
 - ▶ ce categorii de erori sunt mai frecvente în TA;
 - ▶ ce categorii de erori sunt mai frecvente în subtitrare.
3. Materialul folosit pentru subtitrarea automată este format din mai multe videoclipuri de scurtă durată care aparțin unor genuri diferite (documentar, cântec, poezie, umor, rețetă de bucătărie, buletin meteo, indicații GPS etc.)
4. Gradul de dificultate nu este mare în prima etapă (acesta va crește în etapele următoare de analiză)
5. Se folosesc toate programele de subtitrare automată identificate fără însă a le compara între ele (momentan nu se face comparație de software).

SUBTITRAREA AUTOMATĂ

- ▶ Utilă;
- ▶ Rapidă;
- ▶ Ieftină;
- ▶ Accesibilă;
- ▶ Permite înțelegerea filmului atunci când:
 - ▶ conținutul audio sau calitatea video este slabă;
 - ▶ persoane cu deficiențe de auz;
 - ▶ limba sursă a filmului nu se înțelege (barieră lingvistică);
 - ▶ crește vizibilitatea unui videoclip;
 - ▶ permite urmărirea unui videoclip în medii în care păstrarea liniștii este obligatorie.

TEHNOLOGIE:



PLATFORMELE DE SUBTITRARE AUTOMATĂ...

Rețele de socializare

1. YouTube Automatic Caption
2. Facebook's Automatic Caption Feature
3. Instagram Automatic Captions

Generatoare de subtitrări

4. **Media.io** - <https://www.media.io>**Diy captions** - <https://www.diycaptions.com/?ref=cybrhome>
5. **Kapwing Auto-Subtitle** - <https://www.kapwing.com/>
6. **Otter.ai** - <https://otter.ai/>
7. **Animaker** - <https://app.animaker.com/subtitle/9to6cfzZJklokFyA>
8. **Veed.io** - <https://www.veed.io/edit/431b2795-3340-49af-afec-e1d3efaeed29/subtitles/translate>
9. **Flixier** - <https://flixier.com/>
10. **HappyScribe** - <https://www.happyscribe.com/>
11. **Maestrasuite** – <https://maestrasuite.com/>
12. **Transkriptor** - <https://transkriptor.com/>
14. **Wearenova** - <https://wearenova.ai/nova-tools/au>
15. **Flexclip** - <https://www.flexclip.com/tools/auto-subtitle/tomatic-subtitles/>
16. **SubtitleBee** - <https://subtitlebee.com/>
17. **Simplified** - <https://simplified.com/video-editor/auto-subtitle-generator>
18. **Filmora** - <https://filmora.wondershare.com/video-editing-tips/best-automatic-subtitle-generators.html>
19. **EasySsub** - <https://easysub.com/>
20. **ASG** - <https://www.autosubgen.com/>
21. **Zubtitle** - <https://zubtitle.com/?via=eddy>
22. **CapCut** - <https://www.capcut.com/>
23. **Checksub** - <https://www.checksub.com/>
24. **Rev** – <https://www.rev.com/checkout/captionfiles>
25. **wavel.ai** - <https://wavel.ai/studio/auto-subtitles-generator/>
26. **Subtitle video** - <https://www.subtitlevideo.com/auto-subtitle-generator-online.php>
27. **AudioType** - <https://www.audiotype.org/subtitles/>
28. **Wave.video** - <https://wave.video/video-captioning>
29. **Hei.io** - <https://www.hei.io/subtitle-generator>
30. **Clideo** - <https://clideo.com/add-subtitles-to-video>

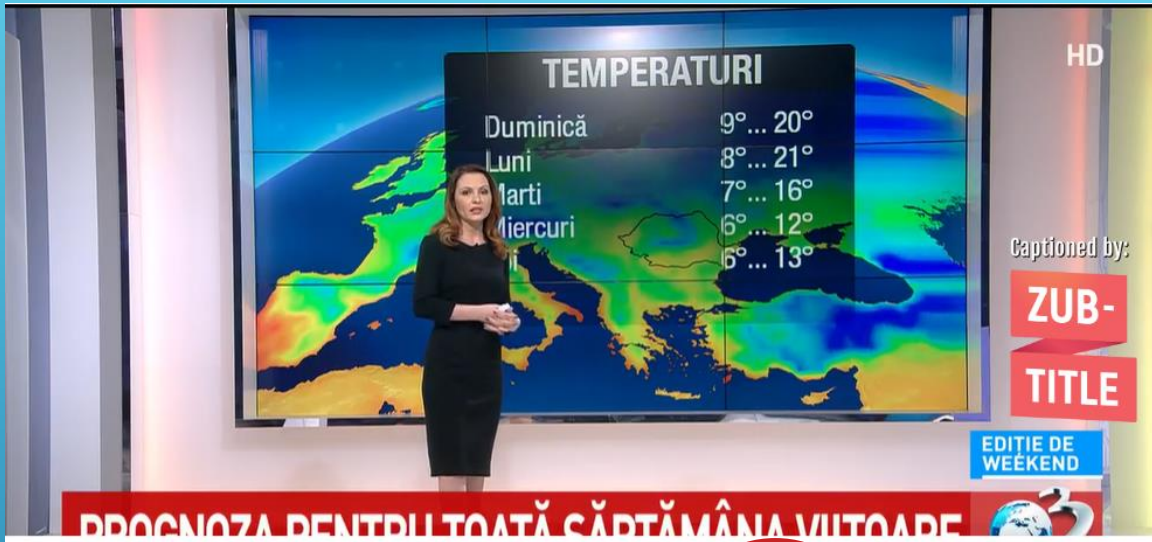
ETAPA I: TRANSCRIEREA AUTOMATĂ

Transcrierea este îngreunată de:

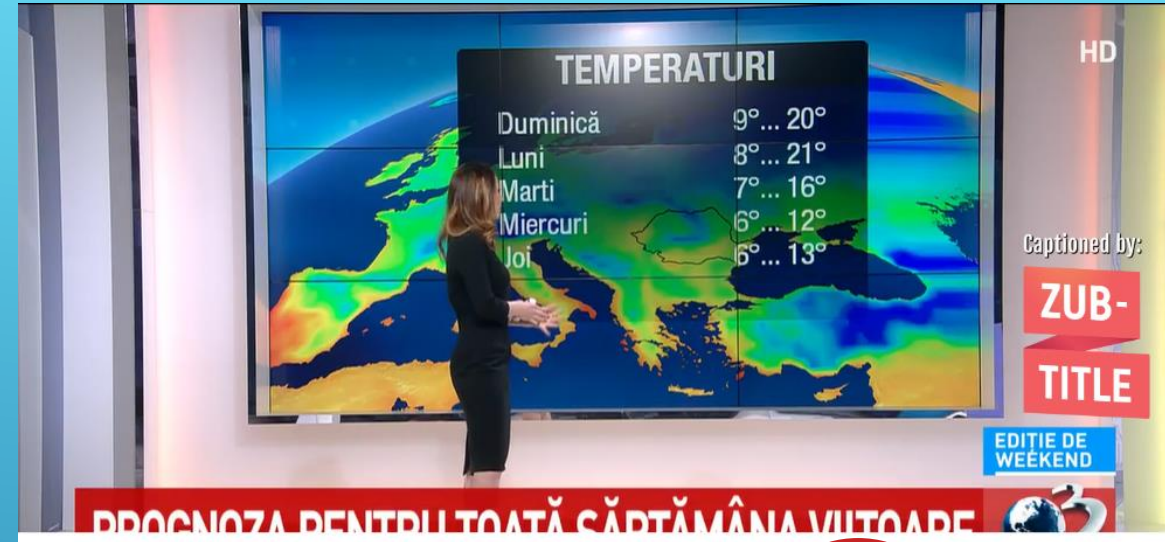
- ▶ Calitatea materialului video
- ▶ Videoclip multilingv
- ▶ Mulți vorbitori
- ▶ Pronunție regională/neclară/arhaică
- ▶ Coloana sonoră (zgomot/muzică)
- ▶ Viteza de vorbire

Erorile de transcriere observate:

- ▶ Transcriere greșită (omofonie)
- ▶ Textul neînțeles nu se transcrie
- ▶ Textul neînțeles este marcat prin [...]
- ▶ Semne de ortografie și punctuație folosite incorect în text
- ▶ Majuscule/Minuscule incorecte,
- ▶ Segmentarea paragrafelor greșită



temperaturile nu vor mai trece de 12 13 grade Celsius



și va fi în regiunile intracarpatiche acolo Unde vor fi valori



EN – The Hobbit. An Unexpected Journey (2012)

For over the Misty mountains cold.
To dungeons deep.
And caravans [...].
We must do it.
Break of day.
To find all.
For God.
The Pines were roaring.
Ohh.
Bowings where morning.
In the night.
The fire was red.
Yeah, it flaming spread.
Trees like torture.
With love.

52 de cuvinte - 13 erori de transcriere

Far over the misty mountains cold
To dungeons deep
and caverns old
We must away ere
break of day
To find our long-forgotten gold

The pines were roaring
on the height
The winds were moaning
in the night
The fire was red,
it flaming spread
The trees like torches
blazed with light

POEZIE PENTRU COPII

RO

Cățeluș cu părul creț - Youtube [Catelus cu parul cret](#)  [poeziile copilariei](#)

Ținuți cu părul creț,
fură rața din coteț
el se jură că nu fură
și l-am prins cu rața în gură
și cu o în buzunar.
Sfatul popular
mă duc fost odată
și am căzut cunsum baltă.

41 de cuvinte - 10 erori de transcriere

Adrian Pinteau - „Glossă” de Mihai Eminescu (@Arhiva TVR)

Vreme trece. Vreme vine.

Toate-s vechi. Și **nou**. Toate.

Ce e rău și ce e bine tu **te întreagă**? Și socoate.

Nu spera. Și nu ai teamă. Ce e val. Ca valul. **Treci**.

De **te Ndeamnă** de te cheamă. Tu rămâi? La toate. Rece.

Multe trec pe dinainte în auz. Ne sună multe. Cine ține toate minte și ar sta să le asculte. Tu.

Așa să te deoparte. Regăsindu-te. Pe tine. Când cu zgomote deșarte vreme, trece vreme vin.

Nici încline a ei limbă recea cumpăna gândirii înspre clipa ce se schimbă pentru masca fericirii

ce din moartea ei se naște și o clipă ține. Poate.

Pentru cine o cunoaște, **toate sunt vechi**.

Și **nu**? **Toate**.

Privitor. Ca la teatru. Tu în lume să **te nchipui**. Joace unul și pe patru totuși tu **ghici vei chipu**

i și de plânge de se ceartă tu în colț? Petreci în tine și înveți din a lor artă ce e rău?

Și ce e bine?

Viitorul. Și trecutul. Sunt a filei două fețe **vedem** capăt începutul, cine știe să le **înveți**?

Tot ce a fost ori o să fie. În prezent. Le avem pe toate.

Dar. Dar lor zădărnice te întreabă.

Și socoate. Căci acelorași mijloace se supun câte există și de mii de ani încoace, lumea ei veselă

și tristă, alte măști, aceeași piesă, alte guri. Aceeași gamă. Amăgit atât de adese. Nu spera.

Și. Nu ai teamă.

Nu spera când vezi mișeeii la izbândă că fac punte **între cele tărâmurii** ai fi cu stea în frunte teamă

N-ai **căta vor** iarăși între dânșii să se plece. Nu te prinde lor. **Tovarăși**.

Ceva. Canalul.

Treci.

Cum cântec de sirenă lumea întinde **lucy Mlj** ca să schimbe actorii în scenă **de** momeste în vârteje

tu. Alături de strecoară. Nu băga nici chiar de seamă. Din Cărarea ta. Afară de **tenda**. De te cheamă.

Pe dating? Să ferii în laturi de hulesc. Să taci din gură. Ce mai vrei cu a tale sfaturi? Dacă știi

a lor măsură zică, toți ce vor să zică? **Pleacă în lume. Cine o treci?** Ca să nu îndrăgești. Nimic.

Tu rămâi e? La toate. Rece.

Rămâi la toate rece de te ndeamnă. De te cheamă. Ce e val ca valul trece. Nu spera și nu ai teamă.

Te întreabă și socoate, ce e rău? Și ce e bine? Toate-s vechi. Și **nu** toate. Vreme trece.

Vreme.

Vine.

YOUTUBE - THE BEST EXPLANATION OF PI

This is pi. But how do we calculate it? First, we'll draw a perfect circle and then its diameter. If we take the diameter to equal one unit, then measuring the circumference will give us pi.

And in school, we learned that pi is equal to the circumference over diameter. So we'll take a look at the circle again, and we have a diameter of one, and measuring the circumference is 3.14,159 and so on, which is equal to pi.



ETAPA II: TRADUCEREA AUTOMATĂ

- ▶ Probleme de traducere în cazul textelor care:
 - ▶ cuprind terminologie,
 - ▶ versuri,
 - ▶ jocuri de cuvinte,
 - ▶ Umor,
- ▶ Filme cu conținut multilingv,
- ▶ Alegerea registrului corect, a stilului (expresivitate, nivel de specializare, grad de formalizare),
- ▶ Înțelegerea și adaptarea culturilor,
- ▶ Evoluția limbii:
 - ▶ Morfologie
 - ▶ Sintaxă
 - ▶ Limba veche (textele istorice)
 - ▶ Varietățile lingvistice (EN-UK, EN-US)
- ▶ Polisemie
- ▶ Neologisme
- ▶ Omonime
- ▶ Schimbări semantice – cloud, catfish, viral

TA nu face proofreading

TA nu acoperă toate limbile

TA nu poate traduce toate tipurile de texte (literatură, marketing, publicitate)

TA nu are acuratețe în traducere

TA va genera mereu același text (nu oferă consiliere, alternative, expertiză, nu se implică în traducere)

TA prezintă probleme de confidențialitate și de securitate: textele se stochează și se procesează de către terți.

TA produce traduceri inegale - pentru unele limbi calitatea este mai bună pentru altele este mai slabă

TA nu poate traduce cuvintele fără echivalent într-o altă limbă

TA nu poate identifica termenii noi

TA nu va înlocui traducătorul

RO – The Hobbit. An Unexpected Journey (2012)

Căci peste munții Cețoși sunt reci.
La temnițe adânci.
și rulote.
Trebuie să o facem.
Pauză de zi.
Pentru a găsi toate Pentru Dumnezeu.

Pinii au fost hohotitor.
Oh.
Bowings în cazul în care dimineața.
Noaptea.
Amenda era roșie.
Da, s-a răspândit în flăcări.
Copaci ca tortura.
Cu dragoste.

Far over the misty mountains cold
To dungeons deep
and caverns old
We must away ere
break of day
To find our long-forgotten gold

The pines were roaring
on the height
The winds were moaning
in the night
The fire was red,
it flaming spread
The trees like torches
blazed with light

EN - Aurel Tămaș, Viață trecătoare

AA.

li MORII puts me in
the pit world and world,
those at my head EO,
market world and world 9
let me die weigh our life.
Transient after forgetting dear world in.
Hi.

I don't know heaven shore
AM and world, or IATA
as I expected M and ara world yes,
here water in life.
By the way, we
liked the world, dear world.

De-oi muri m-or pune-n groapă
Lume și iar lume
Și la cap mi-or pune-o piatră
Lume și iar lume
Două zile m-or cânta
Viață trecătoare
Dup-acee m-or uita
Lume dragă lume.

Nu știu raiu' m-a lua
Lume și iar lume
Or iadu' m-a aștepta
Lume și iar lume
Da' aicea pe Pământ
Viață trecătoare
Am făcut ce mi-o plăcut
Lume dragă lume.

CUVINTE CARE NU AU ECHIVALENT

Still have. One **longing**.
In the stillness of the evening.
Let me die.
At the edge of the sea.

I am either a smooth sleep.
And the **code** is close.

The **talanga** entered.
Cold evening wind.

Lucifers
What a spring from the shadow of **cetini**.

Mai am un singur **dor**
În liniștea serii
Să mă lăsați să mor
La marginea mării.

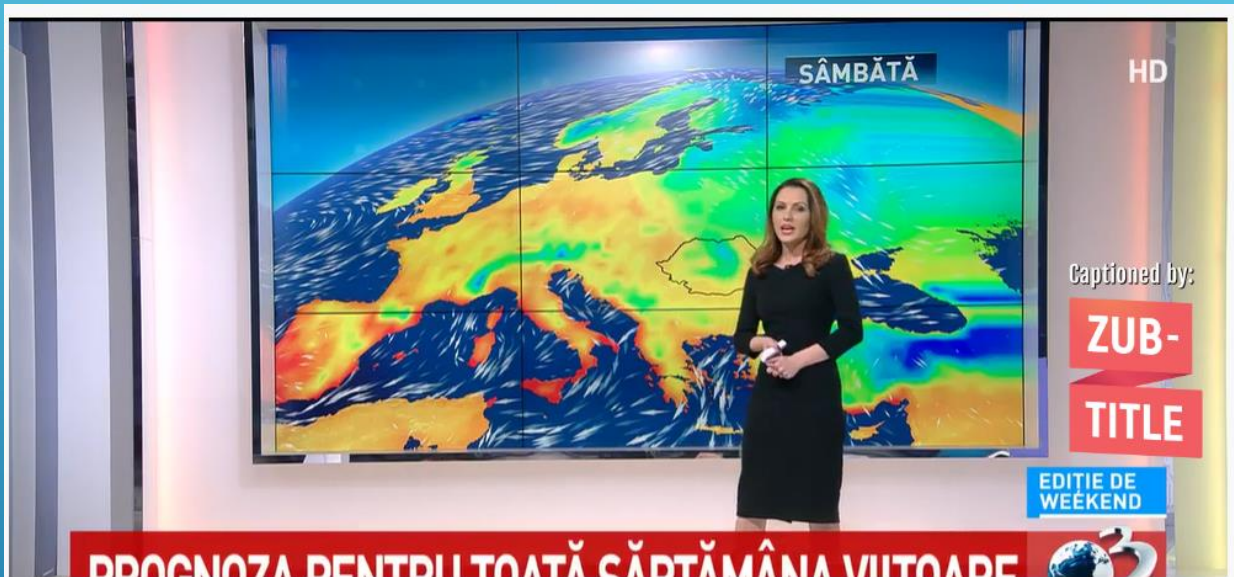
Să-mi fie somnul lin
Și **codrul** aproape,

Pătrunză **talanga**
Al serii rece vânt,

Luceferi, ce răsar
Din umbră de **cetini**,

ETAPA III: SUBTITRARE AUTOMATĂ

- ▶ Două tipuri de subtitrare automată: intralingvistică, interlingvistică;
- ▶ Există o mare varietate de subtitrări (**30 de tipuri diferite de subtitrări în Bartoll 2008: 270-291**);
- ▶ Subtitrări suprapuse peste alte texte afișate pe ecran;
- ▶ Dacă programul nu înțelege, nu subtitrează;
- ▶ Elementele verbale care se văd în film nu se subtitrează;
- ▶ Punctuație incorectă;
- ▶ Nu poate realiza reducția subtitrărilor;
- ▶ Tendința de a subtitra cuvinte inexistente din coloana sonoră a filmului;
- ▶ Probleme formale în subtitrare:
 - ▶ **Timp**: durată minimă/maximă, spațiul între subtitrări, sincronizare, ritm de lectură.
 - ▶ **Spațiu**: nr. minim/maxim de caractere/rând, număr de rânduri, segmentare



prin jumătatea de Sud a țării însă de luni vremea



temperaturii în marți cuprinse
între 7 și 15 grade urmând ca de

#	Start time	End time	Duration	Text
4	00:00:24,520	00:00:28,430	3,910	ATP is released by the mitochondria so cells can use it.
5	00:00:30,240	00:00:35,112	4,872	Mitochondria consists of two membranes, an outer membrane separating it
6	00:00:35,122	00:00:40,620	5,498	from the cytosol and an inner membrane surrounding the so-called matrix.
7	00:00:41,730	00:00:45,820	4,090	The area between these membranes is called the intermembrane space.
8	00:00:47,830	00:00:51,841	4,011	ATP is generated at the inner membrane of mitochondria by
9	00:00:51,851	00:00:58,951	7,100	an efficient mechanism called oxidative phosphorylation involving sever
10	00:01:03,830	00:01:08,780	4,950	Nutrients provide high energy electrons in the form of NADH,
11	00:01:08,880	00:01:11,983	3,103	which are used by the protein complexes to pump protons
12	00:01:11,993	00:01:15,111	3,118	from the matrix to the inter membrane space.
13	00:01:16,810	00:01:21,411	4,601	This continuous pumping creates a proton gradient where the positively
14	00:01:21,421	00:01:25,235	3,814	charged protons are attracted to the more negative matrix.



Start time: 00:00:51,851 Duration: 7,100 Chars/sec: 14,23

Text: **an efficient mechanism called oxidative phosphorylation involving several membrane protein complexes.**

Buttons: Unbreak, Auto br, Split line!

Single line length: 39/62 Total length: 101

Translate Create Adjust

Insert new subtitle at video pos

Play from just before text

Go to sub position and pause

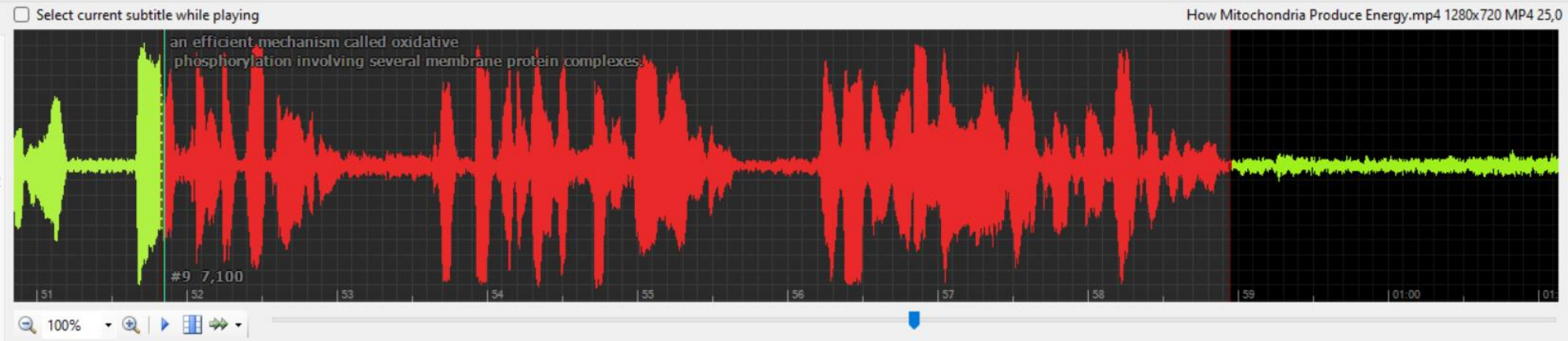
Set start time

Set end time

<< 0,500 >>

<< 5,000 >>

Video position: 00:00:51,851



CONCLUZII

- ▶ Există o mare varietate de platforme de subtitrare automată;
- ▶ Acestea sunt utile pentru publicul larg deoarece permit accesul la un conținut lingvistic aproximativ;
- ▶ Orice element fonetic poate afecta transcrierea;
- ▶ În traducerea automată găsim mult mai multe erori decât în transcriere;
- ▶ Textele specializate sunt mult mai ușor de tradus decât textele literare;
- ▶ Programul, de cele mai multe ori nu respectă limitările formale specifice traducerii audiovizuale;