



# What Is the Productivity Gain in Machine Translation of Subtitles?

Martin Volk - University of Zurich, Switzerland

Mark Fishel - TextShuttle, Switzerland

Lindsay Bywood - VSI and Imperial College, UK

Yota Georgakopoulou - Deluxe Digital Studios, UK





# SUMAT

An Online Service for SUBtitling by MACHine Translation

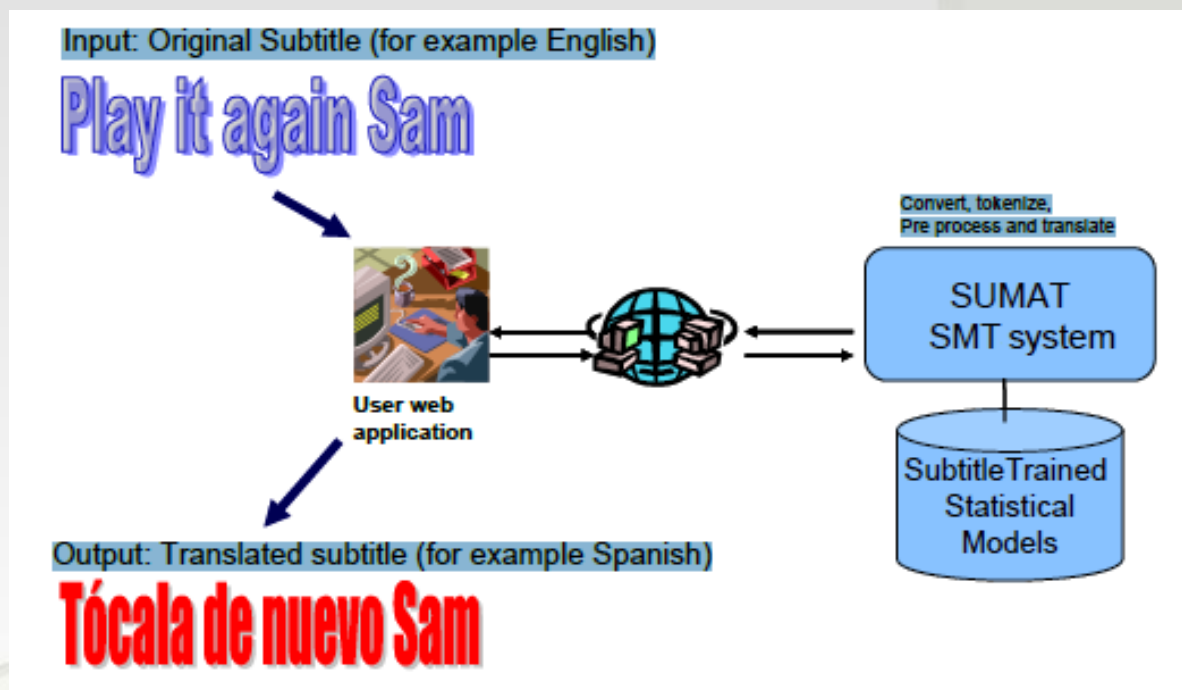
<http://www.sumat-project.eu/>

Project execution: From 01/04/2011 to 31/03/2014





## SUMAT service concept representation



Language pairs:

English<>Spanish

English<>French

English<>German

English<>Swedish

English<>Portuguese

English<>Dutch

Serbian<>Slovenian



## SUMAT EVALUATION

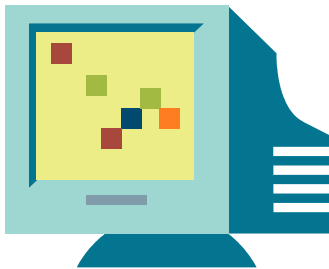
CASE STUDY: JULY 2012 - OCTOBER 2012 (COMPLETE)

1<sup>st</sup> ROUND: FEBRUARY 2013 - JUNE 2013

2<sup>nd</sup> ROUND: SEPTEMBER 2013 - FEBRUARY 2014

### Automatic evaluation

BLEU  
Levenshtein 5



### Human evaluation

Calculation of post-editing  
efficiency





## SUMAT EVALUATION - CASE STUDY

**LANGUAGE PAIR:** English > Swedish

**MATERIAL USED:** 4 videos (1 scripted, 1 semi-scripted, 2 unscripted), plus 1 video for benchmarking (unscripted)

<b>VIDEOS USED</b>	<b>Subtitle number</b>	<b>Runtime</b>	<b>Genre/domain</b>	<b>Difficulty level</b>
Benchmark File	702 subs	45 mins	unscripted / daily life	standard audio
File 1	1218 subs / doing 567 subs only	120 mins / doing 65 mins only	unscripted / documentary	standard/difficult audio
File 2	754 subs	46 mins	unscripted / daily life	standard audio
File 3	668 subs	46 mins	semi-scripted / comedy	standard audio
File 4	736 subs	93 mins	scripted / feature / drama	standard audio



## SUMAT EVALUATION - CASE STUDY

### SUBTLERS PERFORMING POST-EDITING

Two from DDS, two from VSI - all skill Level 1 (scale 1-3)

**INSTRUCTIONS:** Post edit the machine translated files to your usual quality standard.

DDS: Fairly detailed briefing with entire team took place.  
VSI: No significant briefing took place.

**PE CONDITIONS RECORDED:** Any factor that could affect the PE process was noted

- Issues with videos
- Subtitling software used
- PE training



## SUMAT EVALUATION - CASE STUDY

Examples: No post-editing effort required

### English original

From now on, man,  
you got to pay for what you want free.

### Swedish MT output

Från och med nu måste du betala för dig.

### Post-edited Swedish

Från och med nu måste du betala för dig.



## SUMAT EVALUATION - CASE STUDY

Examples: Little post-editing effort required

### English original

I've had enough. Tryin' to impress me with balloon animals.

- How far are we from the hospital, man?
- [Driver] Ten minutes.

### Swedish MT output

Jag har fått nog. Försöker imponera på mig med ballongdjur.

- Hur långt är vi från sjukhuset?
- [förare] tio minuter.

### Post-edited Swedish

Jag har fått nog. Han försöker imponera på mig med ballongdjur.

- Hur långt är vi från sjukhuset?
- [Förare] Tio minuter.





## SUMAT EVALUATION - CASE STUDY

Examples: Significant post-editing effort required

### English original

Like I said, someone  
calling you their old lady.

Because they're a bunch  
of damn fools.

### Swedish MT output

Som jag sa,  
nån ringa sina gamla dam.

De är ett gäng jävla idioter.

### Back translation

Like I said, someone  
phoning their old ladies.

They're a bunch of fucking idiots.

### Post-edited Swedish

Som sagt,  
om nån kallar en "min lilla gumma",

De är ett gäng dårar.



## SUMAT EVALUATION - CASE STUDY

Examples: Significant post-editing effort required

English original	
The reactions I got in the market stalls with the fishermen,	being exhibited in one of the most important galleries of the world,
Swedish MT output	
Reaktioner på marknaden toaletter med fiskarna,	att visa upp er i en av de viktigaste gallerier av världen,
Back translation	
Reactions on the market toilets with the fishermen...	to show yourself in one of the most important galleries from the world...
Post-edited Swedish	
Reaktionen hos fiskarna på marknaden...	som visas upp på ett av världens viktigaste gallerier,



## SUMAT EVALUATION - CASE STUDY FINDINGS

<b>Material</b>	<b>Nature</b>	<b>Avg productivity gain/loss(-)</b>	<b>Multiref BLEU</b>	<b>Multiref Lev5</b>	<b>Multiref avg lev-dist</b>
File 1	unscripted	-49.9% ... 14.7%	57.3	35.27%	9.6
File 2	unscripted	-17.0% ... 23.3%	66.9	51.99%	7.2
File 3	semi-scripted	-39.3% ... 19.2%	70.0	58.98%	6.1
File 4	scripted	1.3% ... 31.6%	75.3	64.95%	4.9



## SUMAT EVALUATION - WHAT HAVE WE LEARNED?

- Scripted genres perform better than unscripted
- Spending time setting up the evaluation is important
- More difficult to increase productivity when subtitlers are not used to a strict template workflow





## SUMAT EVALUATION - FOOD FOR THOUGHT

THINGS TO THINK ABOUT FOR FULL-SCALE EVALUATION:

- Post-editing conditions: skill level, training, preparation
- External factors: template file, video file, interrupted work
- Choice of benchmark file
- Existing workflow
- Results are likely to vary with language





Thank you!  
Questions?

