



Subtitle translation using TM & MT

Judith Klein, STAR AG

- ▲ Translating subtitles
- ▲ Subtitle project in Transit NXT
- ▲ MT for translating subtitles
- ▲ Subtitle project with MT in Transit NXT

Subtitling

Transferring audio-visual information into written information

▲ Intralingual

Audio and written text in the same language

▲ **Interlingual**

Text is a different language to the original soundtrack

Subtitling – Market development

Increasing demand for multilingual videos due to globalisation and technical opportunities

▲ Entertainment and documentary films

▲ **Company and product videos**

English video with English subtitles



they know there's a more efficient way to communicate.

Features and requirements

Film and spoken language vs written text

- ▲ Time and space restrictions
- ▲ Subtitle sequence: two lines & 37-41 characters per line
- ▲ Linguistic reduction

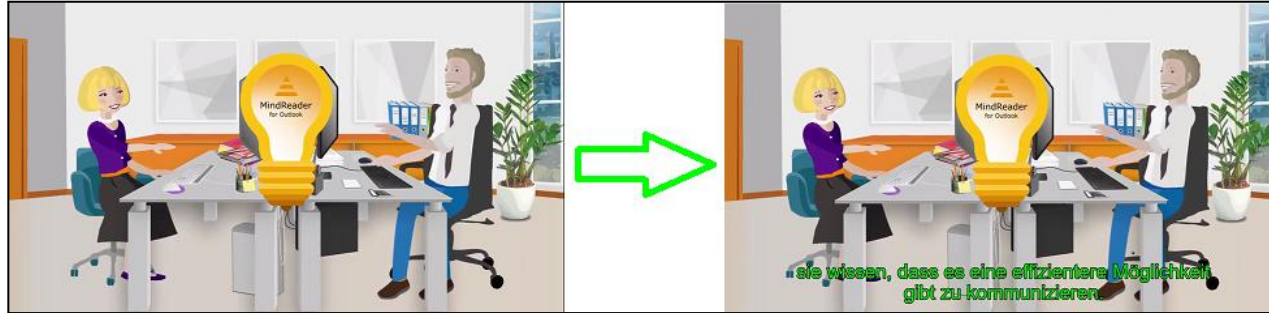
Features and requirements

Film and spoken language vs written text

- ▲ Time and space restrictions
- ▲ Subtitle sequence: two lines & 37-41 characters per line
- ▲ Linguistic reduction
- ▲ Formal, grammatical and stylistic guidelines
- ▲ Target languages are longer than the original language
- ▲ Cultural and business context
- ▲ No misunderstandings

Subtitling methods: (1) Subtitling software

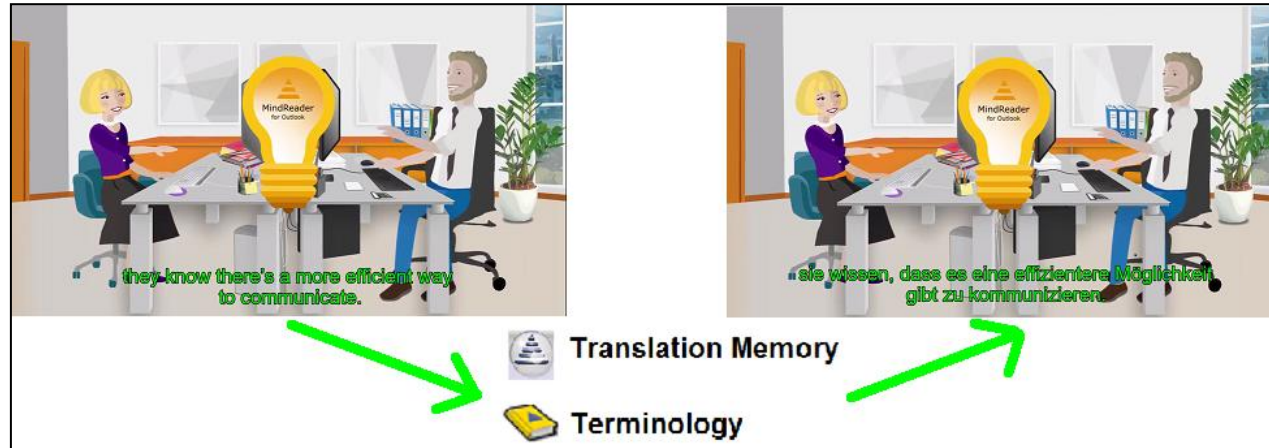
▲ Translate the **original** soundtrack (subtitling software)



- Work in a specific subtitling software
- Target language subtitles created directly in the software
- Translator
 - inserts the target language
 - sets the corresponding time codes

Subtitling methods: (2) Template method

▲ Translate the **subtitles** of the original soundtrack



- Multilingual translation projects (company and product videos)
- Text file with original-language subtitle text and time codes
- No change to the time codes in the target language
- Plan space for longer languages

Excerpt from a subtitle file – characteristics

1

00:00:00,500 --> 00:00:03,069

Mindreader for Outlook ¶

2

00:00:03,069 --> 00:00:09,342

Sophie and Max love working in the ¶
globalized, digital world.

.....

Translating subtitles with CAT tools

- ▲ At the heart of these systems is the translation memory
 - High rate of re-use of existing translations is possible for technical documentation

- ▲ Why use a CAT tool for business videos?
 - Consistent use of company-specific terminology
 - Uniform linguistic style
 - Descriptive language

Translating subtitles in Transit NXT – advantages

▲ Standard translation process

- Import the source-language text file (SRT, VTT, TXT, WebVTT)
- Translate in the specialised, familiar editor
- Export the translation in text format

Translating subtitles in Transit NXT – advantages

▲ Multimedia context

- Audio-visual context → integrated multimedia viewer
- Textual context → reference files in the TM
- Terminological context → graphics, videos

Translating subtitles in Transit NXT – advantages

▲ Length check

- Each subtitle with two lines, each with 40 characters (customizable)
- Further checks for quality assurance: double words, double spaces...
- Terminology check

Example project in Transit NXT, English to German

The screenshot displays the Transit NXT interface with three main panels:

- STAR MR subtitle (English [US]):**

2	MindReader•for•Outlook
4	Sophie•and•Max•love•working•in•the• globalized,•digital•world.
6	Communicating•in•real-time
8	and•setting•up•innovative•projects•is•great.
10	But•if•you•hear•that•constant•ping•of
12	“You’ve•got•mail”
- STAR MR subtitle (German):**

2	+ MindReader•für•Outlook
4	# Sophie•und•Max•arbeiten•gern•in•der•globalisierten,•digitalen•Welt.
6	Communicating•in•real-time
8	and•setting•up•innovative•projects•is•great.
10	But•if•you•hear•that•constant•ping•of
12	+ “Neue•E-Mail”
14	and•you’re•receiving•and•sending•
- Multimedia viewer:** A video player showing a 3D scene with two characters at a desk and a large yellow lightbulb. The video is paused at 00:46.
- Source Fuzzy:**

92%	D:\Transit\NXT\projects\tekom2018\MT_Subtitling_MindReader\ref1\STAR-MR_misc
Ref	Sophie•and•Max•love•working•in•the• in•the•globalized•world.
New	Sophie•and•Max•love•working•in•the• globalized,•digital•world.
Ref	Sophie•und•Max•arbeiten•gern•in• in•der•globalisierten•Welt.
0/1/0/0	Sophie•und•Max•arbeiten•gern•in•der•globalisierten,•digitalen•Welt.
	No•more•fuzzy•matches.[]
- Terminology:**

globalized
DEU globalisiert

Subtitling using a CAT tool – summary

- ▲ Terminology
- ▲ Dual Concordance
- ▲ Dynamic Linking
- ▲ Segment concordance
- ▲ Audio-visual context

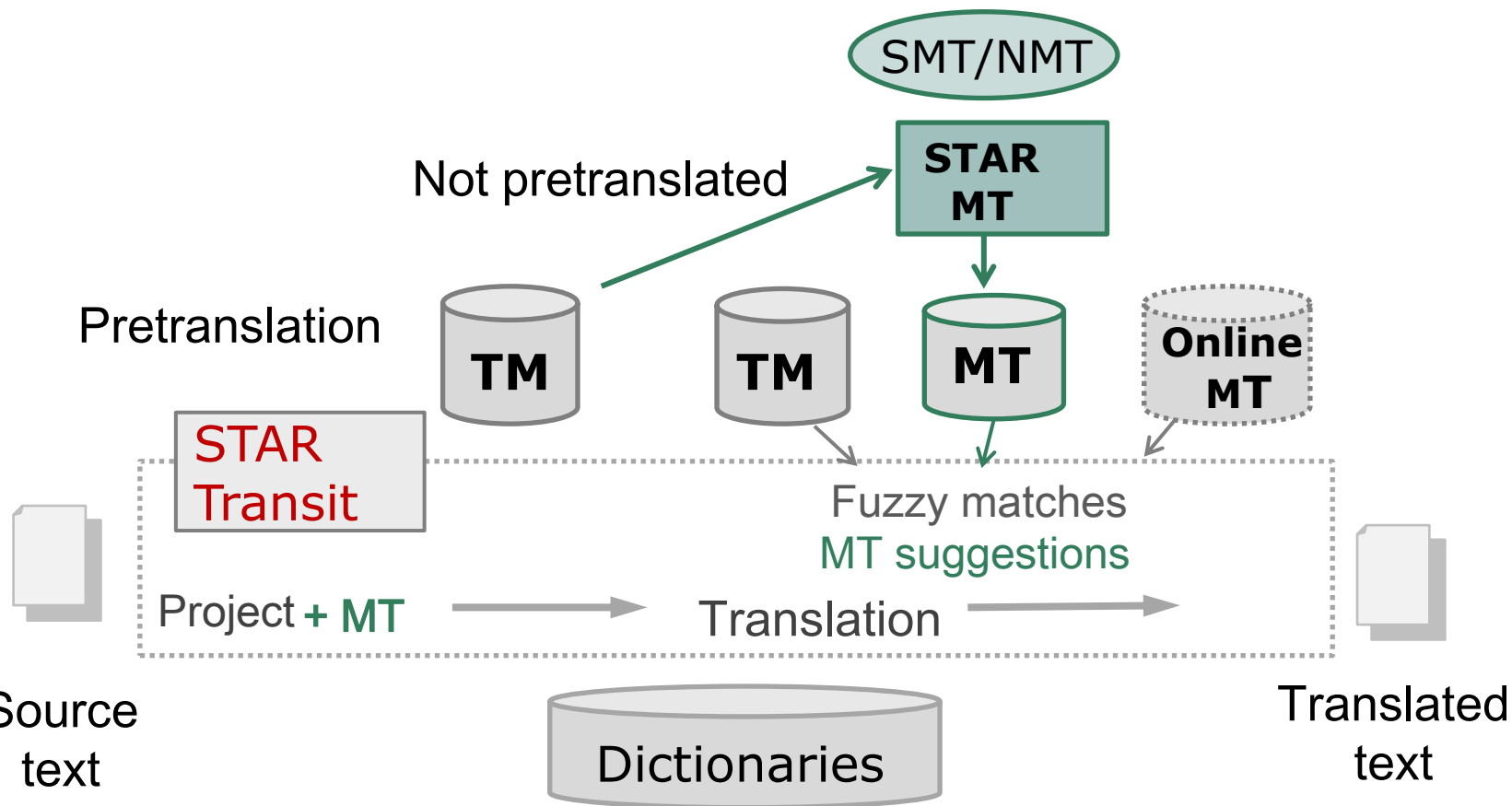
Machine Translation for Translating Subtitles

▲ "Subtitles are well suited for MT"

(cf. Volk 2008, Volk/Sennrich 2011, SUMAT 2014)

- short text units
- little internal complexity (linguistic structure/phenomena)

MT integration in Transit NXT

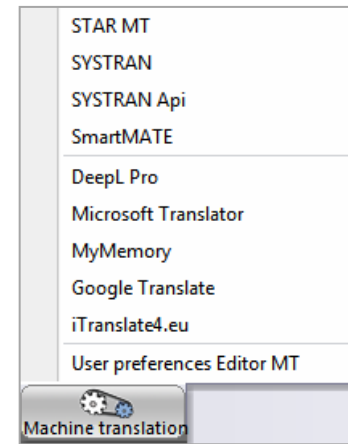


Customer-specific MT systems (e.g. STAR MT)

- ▲ MT suggestions created during import → "Import MT"
- ▲ SMT or NMT engines (size, language, domain)
- ▲ Training data:
 - cleaned, bilingual corpora from customer TMs (e.g. Transit or TMX)
 - technical documents, software localisation, subtitling texts, etc.
 - bilingual word lists (e.g. from TermStar dictionaries)

Interface to online MT systems

▲ Online MT systems → "Editor MT"



162		CBX-MEDIA:
163	+	Combox Multimedia
165		Implausible data were received
167		CBX-ECALL:
168	+	Combox-Notruf
170	+	TRSVC:
171		Overview camera

Find fuzzy matches

Request machine translation

Confirm active segment

Start format check

Machine translation request

For machine translation, the texts and contents that are to be translated are transferred to the machine translation (MT) system, and you will be responsible for this MT system provider. CTAD should provide you with the relevant confidentiality agreements.

you yourself shall be responsible

Note that, when transferring the texts and contents that are to be translated to the MT system or to the MT system provider, you yourself shall be responsible for ensuring compliance with data protection regulations and/or confidentiality agreements.

Machine translation may be a paid service that is offered by the machine translation system provider. Costs may be incurred through the use of machine translation. STAR does not have any influence on the amount or billing of these costs.

Send data to machine translation Skip machine translation

Source Fuzzy	
	Machine Translation (Editor MT)
New	But if you hear that constant ping of
MT	Aber wenn Sie dieses ständige Ping von hören

MT for translators

- ▲ MT not inserted automatically



MT for translators

- ▲ MT not inserted automatically
- ▲ Additional source
 - instead of no match

Source Fuzzy		
		Machine Translation (Import MT)
New		choosing the right wording.
MT		Der Auswahl der richtigen Formulierung.



MT for translators

- ▲ MT not inserted automatically
- ▲ Additional source
 - instead of no match
 - in addition to fuzzy match

Source Fuzzy	
72%	
Ref	 they know there is a better way to communicate.
New	they know there's a more efficient way to communicate.
Ref	sie kennen einen besseren Weg der Kommunikation.
	Machine Translation (Import MT)
New	they know there's a more efficient way to communicate.
MT	Sie wissen, dass es eine höhere Methode ist, um miteinander zu kommunizieren.

MT for translators

- ▲ MT not inserted automatically
- ▲ Additional source
 - instead of no match
 - in addition to fuzzy match
- ▲ Clearly labelled

Source Fuzzy	
72%	
Ref	 they know there is a better way to communicate.
New	they know there's a more efficient way to communicate.
Ref	sie kennen einen besseren Weg der Kommunikation.
	Machine Translation (Import MT)
New	they know there's a more efficient way to communicate.
MT	Sie wissen, dass es eine höhere Methode ist, um miteinander zu kommunizieren.

MT for translators

- ▲ **MT not inserted** automatically
- ▲ **Additional** source
 - instead of no match
 - in addition to fuzzy match
- ▲ Clearly **labelled**
- ▲ **Unknowns** indicated

Source Fuzzy	
	Machine Translation (Import MT)
New	More advantageous for long-series production. ⁽¹⁾
MT	Plus avantageux pour long-series production. ⁽¹⁾

MT for translators

- ▲ MT not inserted automatically
- ▲ Additional source
 - instead of no match
 - in addition to fuzzy match
- ▲ Clearly labelled
- ▲ Unknowns indicated
- ▲ TM-validated MT

Source Fuzzy		
		TM-validated MT
New		Pressure increases too slow when filling reservoir
		Druckanstieg zu langsam bei Füllung des Tanks

Working with MT suggestions

▲ MT suggestions only for specific segments possible:

- if no high fuzzy match

Only for segments with fuzzy matches lower than (%): 80

- for segments of defined length

Only for segments with at least (words): 3

Only for segments with not more than (words): 30

MT for translators – MT position

List MT suggestions in the fuzzy window

Above the best fuzzy match

In the same way as a fuzzy match at (%):

75

List MT suggestions in the fuzzy window

Above the best fuzzy match

In the same way as a fuzzy match at (%):

75

		Machine Translation (Import MT)
		Implausible data were received from control unit "Traction system"
		Wurden nicht plausible Daten empfangen vom Steuergerät Traktionssystem
79%		D:\Transit NXT\projects\MT_Example_AEXTR_1
		Implausible data were received from control unit "Battery".
		Implausible data were received from control unit "Traction system"
		# Vom Steuergerät "Batterie" wurden nicht plausible Daten empfangen.
74%		D:\Transit NXT\projects\MT_Example_AEXTR_1
		Implausible data were received from control unit "Radar sensorics".
		Implausible data were received from control unit "Traction system"
		# Vom Steuergerät "Radarsensoren" wurden nicht plausible Daten empfangen.

79%		D:\Transit NXT\projects\MT_Example_AEXTR_1
		Implausible data were received from control unit "Battery".
		Implausible data were received from control unit "Traction system"
		# Vom Steuergerät "Batterie" wurden nicht plausible Daten empfangen.
		Machine Translation (Import MT)
		Implausible data were received from control unit "Traction system"
		Wurden nicht plausible Daten empfangen vom Steuergerät Traktionssystem
74%		D:\Transit NXT\projects\MT_Example_AEXTR_1
		Implausible data were received from control unit "Radar sensorics".
		Implausible data were received from control unit "Traction system"
		# Vom Steuergerät "Radarsensoren" wurden nicht plausible Daten empfangen.

QA for MT-based translations

- ▲ Markup automatically inserted
- ▲ QA on MT-based translations
- ▲ Statistics on MT-based translations

DEMO – Subtitle project with MT in Transit NXT

- English → German
- Subtitles of STAR product video (software)
- Training data for MT system: TM and generic corpus
- 280,000 segments
- SMT engine

Subtitle project with MT in Transit NXT


TermStar (project dictionaries) STAR MR subtitle.DEU

STAR MR subtitle [English (US)]		STAR MR subtitle [German]	
34	Sophie•and•Max•are•smart•-	34 +	Sophie•und•Max•sind•clever•-
36	they•know•there's•a•more•efficient•way•¶ to•communicate.	36	they•know•there's•a•more•efficient•way•¶ to•communicate.
38	They•decided•to•use•MindReader•¶ for•Microsoft•Outlook	38	They•decided•to•use•MindReader•¶ for•Microsoft•Outlook
40	so•they•can•now•write•as•quickly•as•they• speak.	40	so•they•can•now•write•as•quickly•as•they•speak.

Source Fuzzy

	Machine•Translation•(Import•MT)
New	they•know•there's•a•more•efficient•way•¶ to•communicate.
MT	Sie•wissen•es•ist•ein•effizienter•Weg•¶ zu•kommunizieren.
72%	D:\Transit\NXT\projects\MT_MindR\REF\Ref_x\STAR\MR\subtitle_2017
Ref	they•know•there's•a•better•way•¶ to•communicate.
New	they•know•there's•a•more•efficient•way•¶ to•communicate.
Ref	sie•kennen•einen•besseren•Weg•¶ der•Kommunikation.

Multimedia viewer



Wiedergabe von "STAR MR subtitle": 804 kBit/s 00:46

Terminology

efficient

A effizient

B Syn.: schnell

Dictionary: TransTec1 (TermStar NXT Received)

Pain or gain?

- ▲ MT can be an additional help in a sophisticated translation environment.
- ▲ MT is meant to support translation – more like fuzzy matches, **not like perfect matches.**
- ▲ Understanding "MT style" will make it easier.