



Use MT to Simplify and Speed Up Your Alignment for TM creation

Judith Klein, STAR Group

Judith.Klein@star-group.net

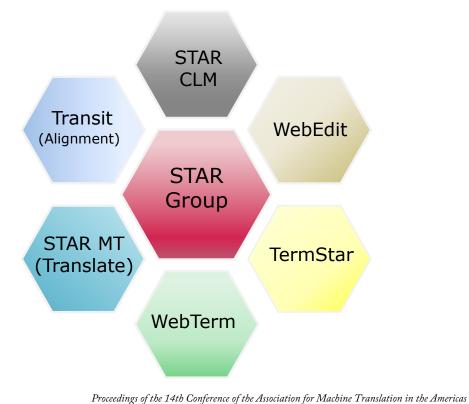
- Motivation
- > TMs for CAT Tools
- MT-based Alignment
- Workflow for MT-Aligned TM
- Conclusion

STAR Group (since 1984)

Translation Services & Translation Technology



STAR's Translation Technology

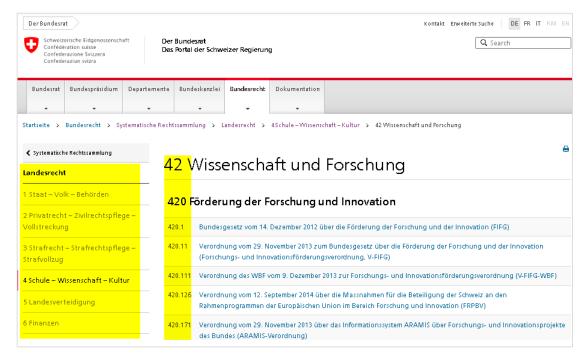


October 6 - 9, 2020, Volume 2: MT User Track

Swiss Federal Administration (Bund)

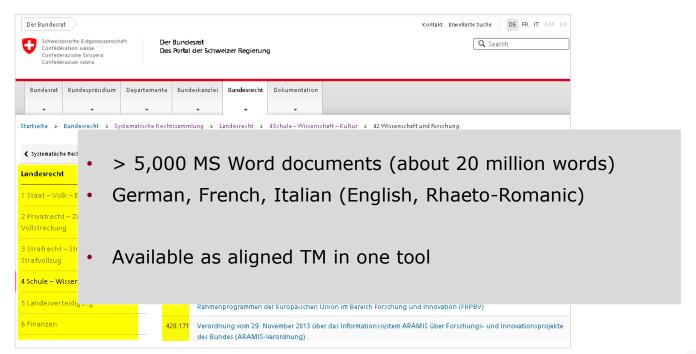
- > 8 departments, 80 offices, 450 language experts
 - Federal Chancellery (BK)
 - Foreign Affairs (EDA)
 - Home Affairs (EDI)
 - Finance (EFD)
 - Justice and Police (EJPD)
 - Environment, Transport, Energy and Communications (UVEK)
 - Defense, Civil Protection and Sport (VBS)
 - Economic Affairs, Education and Research (WBF)
- German, French, Italian, Rhaeto-Romanic, English
- Different tools
- > Huge amounts of language resources (translation pools, termbases)

Systematic Collection of Legislation (SR)



Proceedings of the 14th Conference of the Association for Machine Translation in the Americas October 6 - 9, 2020, Volume 2: MT User Track

Systematic Collection of Legislation (SR)



SR as TM for all Departments

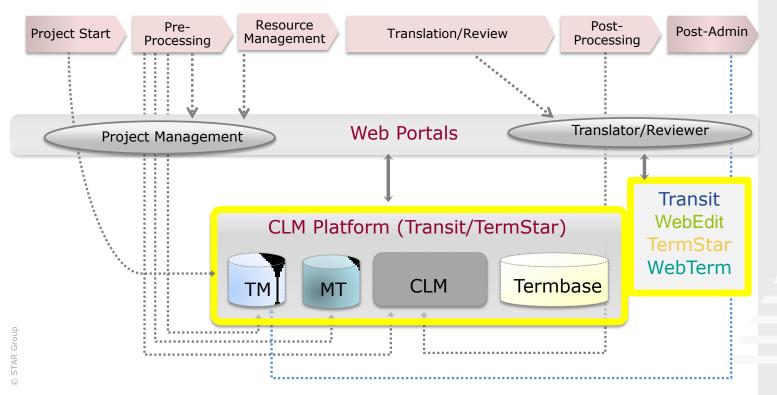
> Aim

- Multi-lingual and multi-directional translation memory
- Access for all departments in the same way
- Use in CAT tool for all search functions

Solution

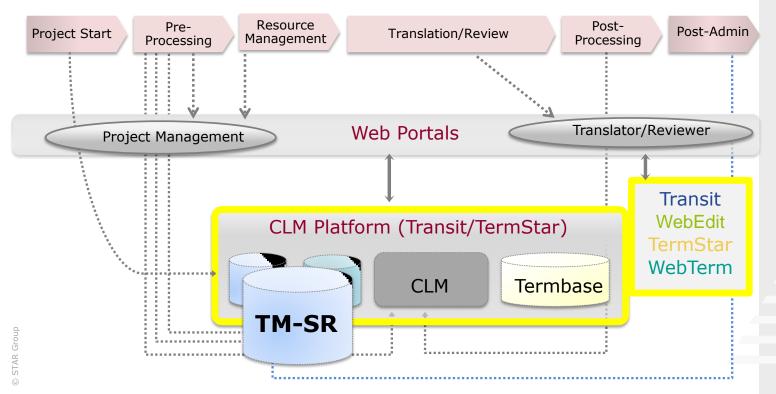
- Standard translation workflow solution
- Fully automated machine alignment
- Fully automated update-workflow every three month

Standard Translation Workflow Solution

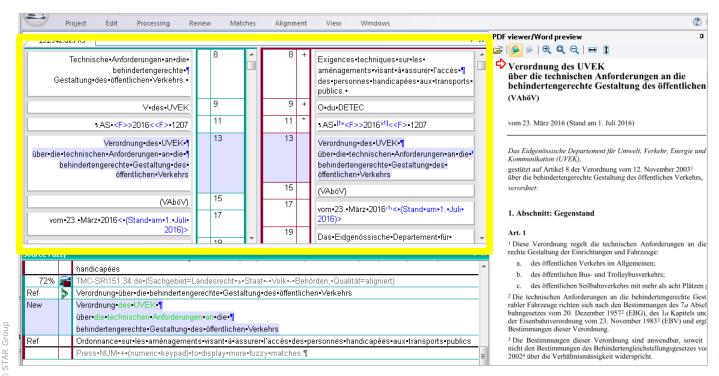


Page 245

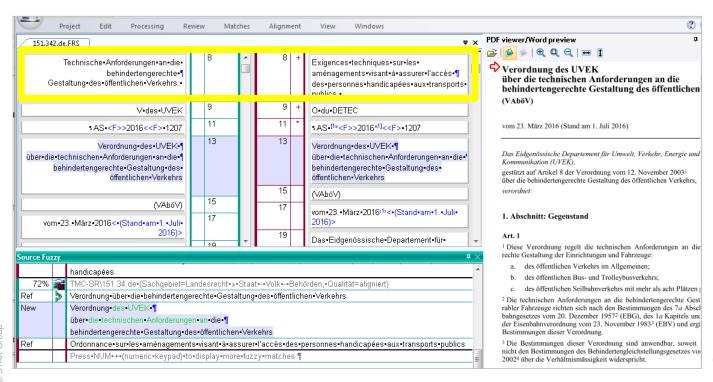
Standard Translation Workflow Solution



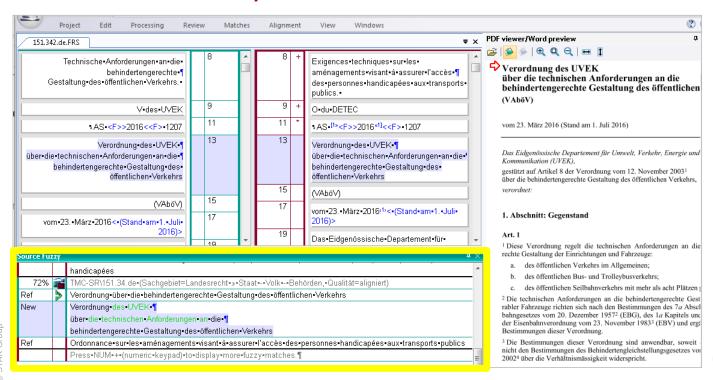
CAT Tool – Segments & Document Context



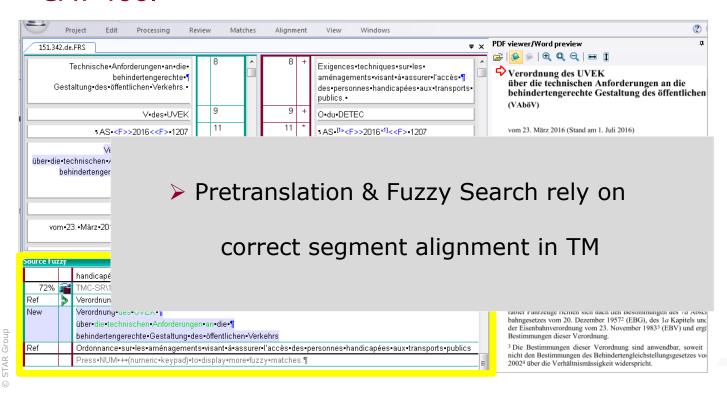
CAT Tool - 100% Matches & Pretranslation



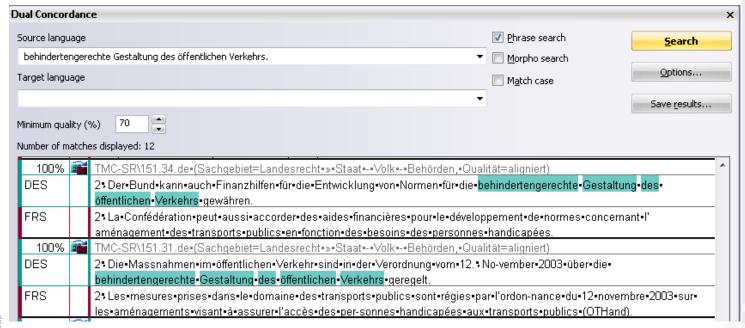
CAT Tool – Fuzzy Matches



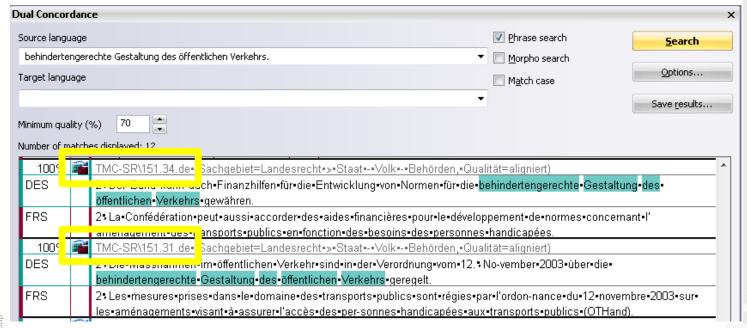
CAT Tool



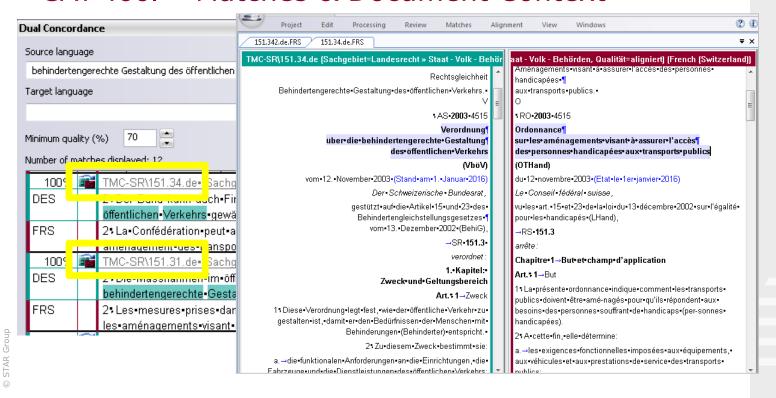
CAT Tool - Concordance Search



CAT Tool - Concordance Search

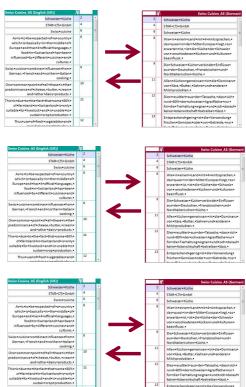


CAT Tool - Matches & Document Context



Alignment





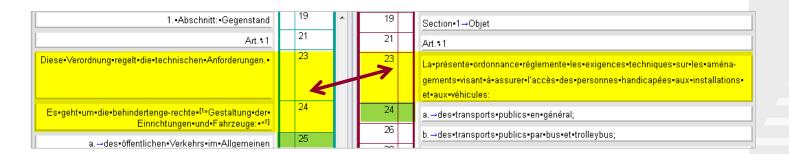
Thus-use-of-fresh-vegetables-and-

Why not MT instead?

- Documents approved by linguists and law professionals
- Specific terminology usage and meaning depending on sub-areas
- Consistent translation required for same/similar sentences
- Look-up text and translation of specific law (e.g. file 151.342)
- > MT often very well but different
- MT can be identical by chance but not reliably so
- Systematic Collection of Legislation must be available as it is!

CAT Tool & Alignment

- Formatting
- Linguistic
- ▶ 1 <> many, 0<> 1 , different order



CAT Tool & Alignment

- Alignment projects with file alignment
- Import and calculate alignment probability
- > Formal (sentence length, numbers, formating, characters,...)
- Lexical (unchanged words, dictionary entries, word lists)

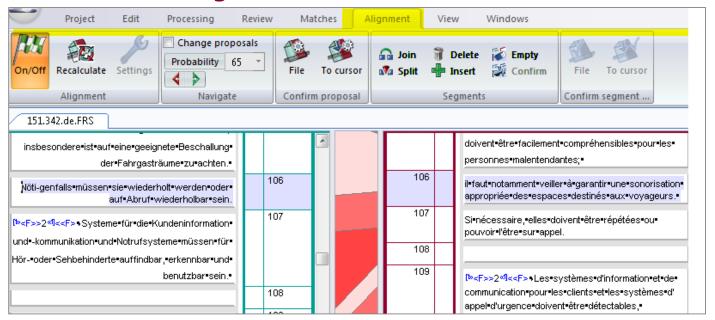
STAR Transit & TM-based Alignment

- Translation Memory segments for alignment
 - Search in TM for segment that is to be aligned
 - Compare translation from TM with alignment candidate
 - Use similarity score as most important value for probability
 - Transit fuzzy algorithms to calculate the similarity
- But no TM available for the Systematic Collection of Legislation

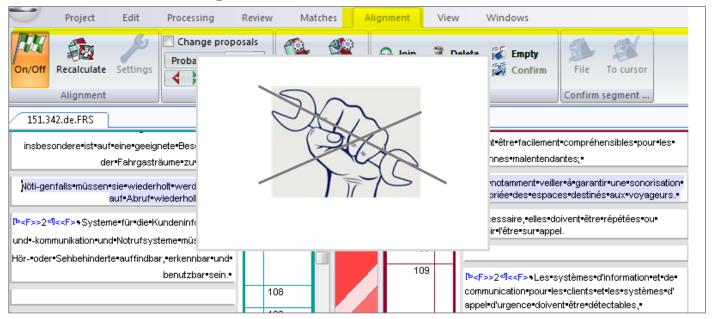
STAR Transit & MT-based Alignment

- Use MT!
- Transit fuzzy algorithm as with TM segments
 - MT-based sentence alignment (Sennrich & Volk, AMTA 2010)
 - BleuAlign used in BiTextor within Paracrawl project
- MT quality per se not so important
- Similarity to decide if two sentences are translation of each other
- MT interfaces available
- MT system selection (DeepL)

Interactive Alignment



Interactive Alignment



Machine Alignment in Transit

- One environment, no separate tool necessary
 - Transit & Alignment tool for project specification
 - Transit MT interface for fast processing



- Import, machine translation, translation comparison, calculation
- Automatic alignment
 - top-down process
 - paragraph tags as anchors
 - 100% similarity as strongest anchors within paragraphs

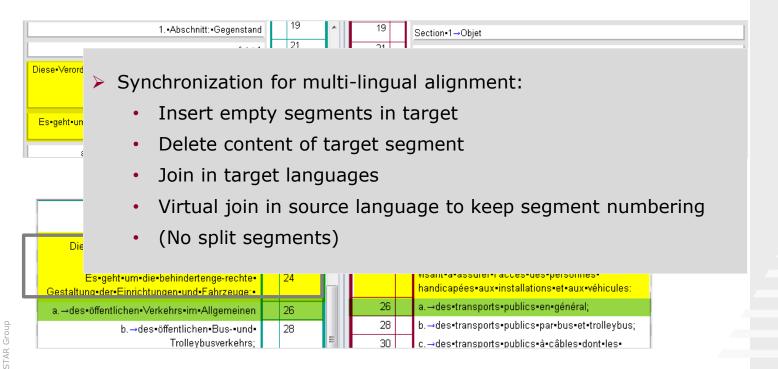


Machine Alignment

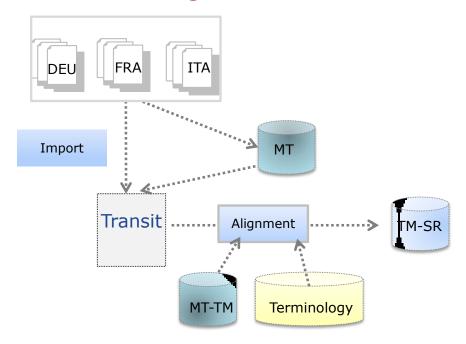
1.•Abschnitt:•Gegenstand Art.•1	19 21	_		19 21	Section•1→Objet Art.•1
Diese-Verordnung-regelt-die-technischen-Anforderungen	23			23	La-présente-ordonnance-réglemente-les-exigences-techniques-sur-les-aména- gements-visant-à-assurer-l'accès-des-personnes-handicapées-aux-installations- et-aux-véhicules:
Es•geht•um•die•behindertenge-rechte• ^[1] *Gestaltung•der• Einrichtungen•und•Fahrzeuge:• ^{«1]}	24	П		24	a. →des•transports•publics•en•général;
a. →des•öffentlichen•Verkehrs•im•Allgemeinen	25		_	20	b.→des•transports•publics•par•bus•et•trolleybus;

1Abschnitt:-Gegenstand		19	^	19	Section•1→Objet
Art.•1		21		21	Art.•1
Diese-Verordnung-regelt-die-technischen- Anforderungen Es-geht-um-die-behindertenge-rechte- Gestaltung-der-Einrichtungen-und-Fahrzeuge:-	П	23 24		23	La-présente-ordonnance-réglemente-les- exigences-techniques-sur-les-aména-gements- visant-à-assurer-l'accès-des-personnes- handicapées-aux-installations-et-aux-véhicules:
a.→des•öffentlichen•Verkehrs•im•Allgemeinen		26		26	a.→des•transports•publics•en•général;
b.→des•öffentlichen•Bus-•und•		28		28	b.→des•transports•publics•par•bus•et•trolleybus;
Trolleybusverkehrs;				30	c.→des•transports•publics•à•câbles•dont•les•

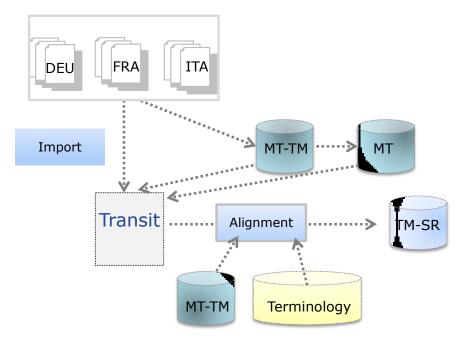
Machine Alignment



Workflow for MT-Aligned TM-SR



Workflow for MT-Aligned TM-SR



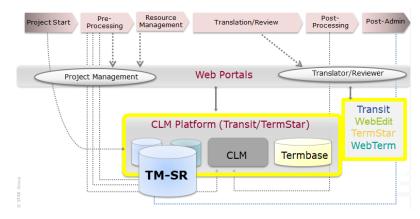
Proceedings of the 14th Conference of the Association for Machine Translation in the Americas October 6 - 9, 2020, Volume 2: MT User Track

Workflow for MT-Aligned TM-SR

- Delta process for next delivery
 - Automatic file comparison
 - Only new and modified documents are processed
 - Deleted files automatically deleted from TM-SR

Conclusion

- Machine alignment integrated Transit feature
- Smooth interaction of Transit, Alignment tool and MT interface
- Random evaluation of machine aligned documents promising
- Use for search functions in CAT tool possible but restricted
- > TM-SR is available for all users in standard translation environment





AMTA(*)20 VIRTUAL 20

Thank you!

Judith Klein, STAR Group

Judith.Klein@star-group.net