

PangeaMT – Start planning your MT adoption for success!

Elia Yuste
PangeaMT, Pangeanic



ANNUAL CONFERENCE 2010

**“Evolution of the Translation Industry -
planning for success”**

Thu, 23rd September 2010

SOAS, University of London, UK



association of translation companies

Agenda

- Insight about (the usefulness of) PangeaMT
- Guidance to implement a PangeaMT translation automation strategy at your LSP company
 - Customisation, domain/client-specific implementations, Post-Editing, advice before and after PangeaMT intake, etc.
- Q&A
- Your benefit as an ATC Conf 2010 Participant
- Appendix: What lies underneath PangeaMT

What is PangeaMT?

Open standards

M O S E S

Going beyond open-source, plain-text only SMT: TMX & XLIFF filters / generators

An innovation-driven machine translation technology conceived as a good-enough qualitative and affordable means to automate translation, speed up and optimise processes, and set you free from user lock-in and extremely dear GILT architectures

I
n
l
i
n
e
r

Wealth of MT services (including consulting & training) → from and for the LSP!

Customised, domain- / client-specific MT solutions (use cases available u.r.)

<http://www.pangea.com.mt/faq/>

PangeaMT's mission & key concepts

- Mission: **Bring democracy and affordability to MT**
 - How? By fostering **open-standard geared translation automation strategies**
- Key concepts
 - PangeaMT is not meant to substitute production environments but to enhance workflows, increasing productivity and cutting down costs
 - PangeaMT's strength is **translation automation & customisation**, heavily relying on our own experience both as international language service vendors and MT researchers and implementers

What we will (and will not) do for you 😊

- We will not provide you with any proprietary architecture for your pre- and post-processing tasks which would make you dependant, but with an MT solution to gradually become fully technology independent.



Are you happy with your current translation environment?

■ YES

→ You may continue to use it to prepare your TMX/XLIFF files for MT input and get them back from PangeaMT as output, ready for post-editing – once more in the same environment.

- This would minimise the familiarisation slope that comes with any new implementation

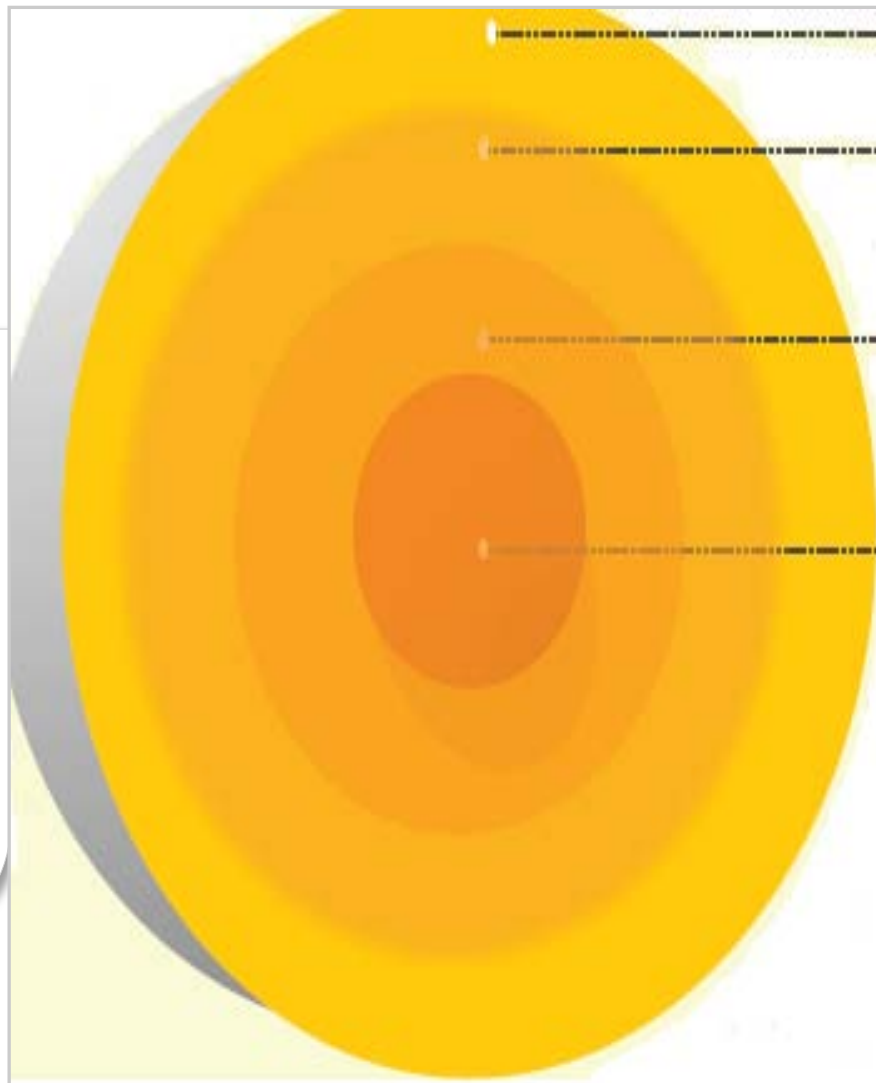
■ NO

→ If you dislike it because you find it for instance too expensive, we can give you some orientation about open-source alternatives that are available in the market nowadays.

- Most CAT systems can import TMs and make use of penalisation features anyway, thus being perfect interfaces for MT post-editing.

Service Workflow

Running centrally on a server, the engine can provide XLIFF / TXT / TMX files, ready for post-editing, and then publishing



Publishing Environment
(DTP, web, etc.)

Post-Editing Environment
(old CAT system or XML-compatible editor)

Service Environment
(intra-/extranet & multiple request handling)

MT Engine Service
(Linux / Cygwin / Web browser)

PangeaMT suggests and helps you implement...

- Proposed workflow based primarily on **MT + Post-Editing (PE)**
 - Post-edited MT output then stored for retraining purposes (stats-driven MT paradigm)
 - A TM-only-based workflow is outdated and entails undesired lock-in effect
- What makes PangeaMT special → open standard format processing & generation capabilities

Customisation – key to success

- Client / domain adaptation
- Data → Not only quantity, but also quality maxime
 - relevant & representative
 - cleanliness
- Pangeanic – a lifelong, proactive member of TAUS TDA
 - helps us offer Data Consultancy services
 - plus other computational linguistics techniques

Post-Editing (PE) & the SMT paradigm → Productivity gains

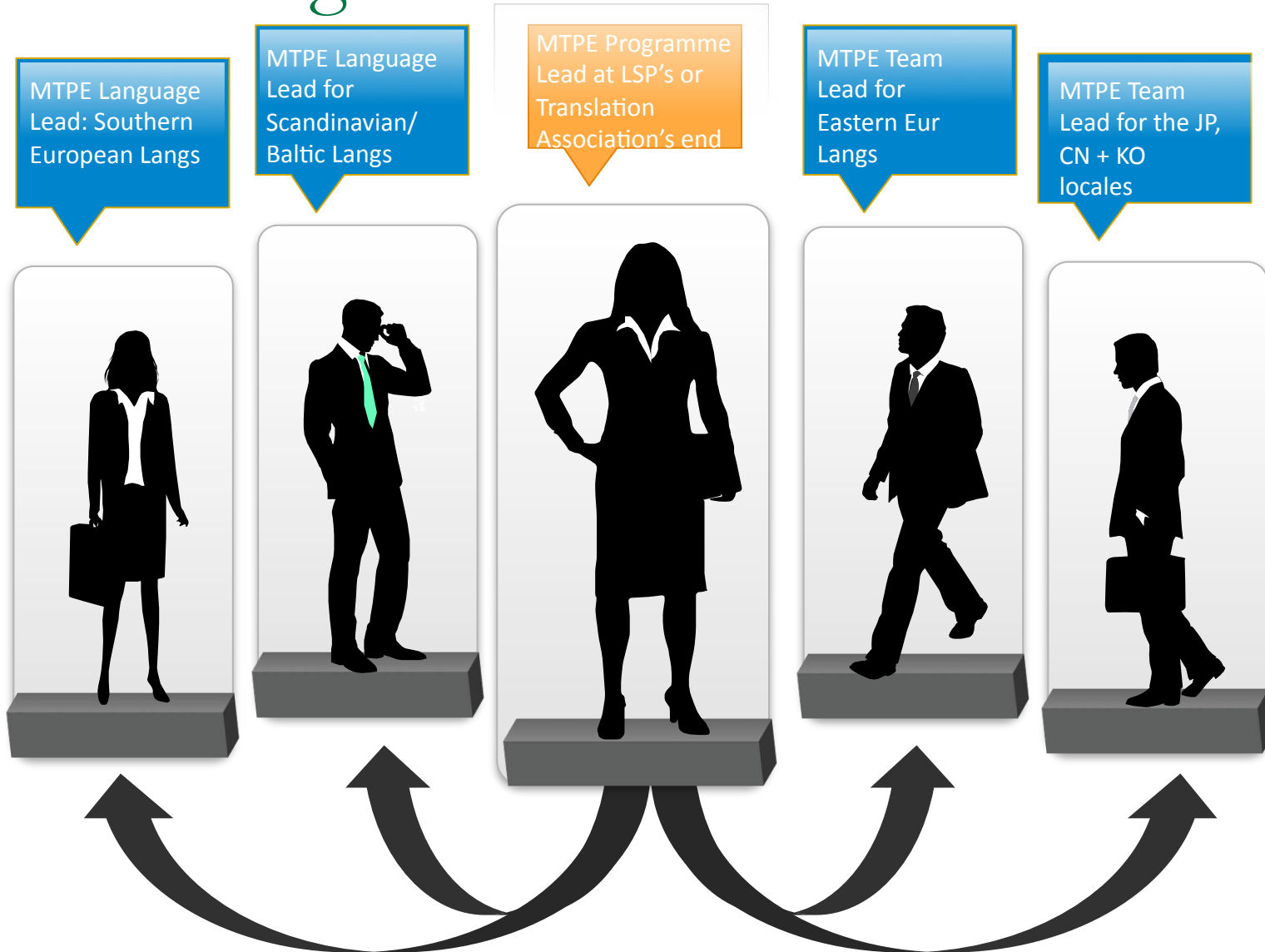
- PE: extremely important in an MT-driven production lifecycle, both to ensure publication quality and to help the MT engine reach *maturity*
 - Use that humanly post-edited MT output to train your customised engine again on a regular basis until you see fit
 - Machine learning leads to better output > gradually less-intensive PE > more productive
 - E.g. presentation by K. Bier (Sybase) and M. Herranz (Pangeanic/PangeaMT) at LW Berlin 2010 → productivity gains ranging from 25 to 300 % depending on content type

PE & Content Leveraging

Procedure to make the best of your MT output

- Adopting MT in your production workflow does not mean to abandon a CAT-like environment but to adjust and exploit it in such way that you can enjoy, if you like, the best of two worlds.
- Set your TM matching options to leverage everything above e.g. 75 % match from the memory and send everything below that to your customised PangeaMT engine in a TMX/XLIFF file.
 - You can post-edit the TMX/XLIFF output in your preferred translation environment again, ideally using one that your linguists are familiarised with

MTPE Programme



Recommendation before adopting a PangeaMT strategy

- Customised and domain-specific PangeaMT solutions makes you stand out of the crowd! → TRUE, but just don't start with a thousand language pairs and domains...
 - Technically speaking, SMT is quick to implement
 - However, you want to adapt to MT (esp. if totally new to it!) and let your team make the most of it
 - Gradual investment in engine development; rapid ROI anyway! → A PILOT highly recommended!
 - What is your key vertical? Client? Language pair?

What to do after PangeaMT intake

- Monitored usage will allow for measuring results systematically and carry out client-specific studies
 - Productivity-related
 - Determine ROI and cost-effectiveness (yearly, cross-annually, per languages, across agency offices – if having more than one)
- Keep in touch with us – help us get better, for your own sake and that of the GLT industry!

PangeaMT *in action*



pangea.com.MT

The Statistical MT Customization Company

E. Yuste

ATC Annual Conf 2010, London, 23rd Sept 2010

http://www.pangea.com.mt/client-area/

Más visitados Comenzar a usar Firefox Últimas noticias

Client Area | pangea.com.mt

pangea.com.MT
The Statistical MT Customization Company

We build your engines

Welcome! » Description » Client Area Contact News FAQs Users » Blog

English ▾

Client Area

User: xxxxxxxxxxxxxxxxxxxxxxx@eu.sony.com

Password: ●●●●●●●●

Validar

News

- Pangeanic welcomes Sony Europe Localization Director
- Pangeanic, first LSP to create SMT division deploying TDA data, TAU\$ Data Association partner highlights
- PangeaMT BDM attends EAMT 2010
- Pangeanic CEO to speak in Localization World



E. Yuste

ATC Annual Conf 2010, London, 23rd Sept 2010



pangea.com.MT

The Statistical MT Customization Company

Plain Text (without tags)

Empty text input field for plain text.

2) 3)

1)

TMX or TXT File Language Pair Engine

 English -> Spanish Electronics

4)

Tag-Optimal Trans-Optimal (beta)

After clicking on Translate, the MT file gets delivered to the client's mailbox in a matter of seconds !

Translation

Empty text input field for translation results.

If you can not use or visualize this page correctly try using Mozilla Firefox



Output extract – *acceptable* quality & right inline tag placement

```

11 <tu creationdate="20100915T125606z" creationid="MT!">
12   <tuv xml:lang="EN-GB">
13     <seg>For example, it can hold a roll of paper for 700 10 x 15 cm (4 x 6 inches) prints, 400 13 x 18 cm (5 x 7
14   </tuv>
15   <tuv xml:lang="ES-EM">
16     <seg>Por ejemplo, puede efectuar racha de papel para 700 10 x 15 cm (4 x 6 pulgadas), imprime 400 13 x 18 cm
17   </tuv>
18 </tu>
19
20 <tu creationdate="20100915T125606z" creationid="MT!">
21   <tuv xml:lang="EN-GB">
22     <seg>High throughput</seg>
23   </tuv>
24   <tuv xml:lang="ES-EM">
25     <seg>Alta capacidad de proceso</seg>
26   </tuv>
27 </tu>
28
29 <tu creationdate="20100915T125606z" creationid="MT!">
30   <tuv xml:lang="EN-GB">
31     <seg>The UP-DR250 can produce up to 450 10 x 15 cm (4 x 6 inch) prints per hour.&lt;br /&gt;&lt;br /&gt;This
32   </tuv>
33   <tuv xml:lang="ES-EM">
34     <seg>El UP-DR250 puede producir hasta 450 10 x 15 cm (4 x 6 pulgadas) por hora. en&lt;br /&gt;&lt;br /&gt;Est
35   </tuv>
36 </tu>
37

```

Special benefits for the LSP company

Deployment of **domain-specific MT technologies** allows for higher production volume and ensures industry vertical specialisation

- Translation automation is the real answer to rate reduction pressure and quicker TAT
- It also contributes to
 - Business dinamisation in existing languages and expansion in other language combinations and potentially, other domains
 - End-client satisfaction

Summing up

We do not sell *existing* software, we believe in pre-existing know-how about MT, techniques, development experience, etc. in order to develop **your** engines. Engines that will not be wrapped in a software version 6, then 7 and 8.5 for which you would have to keep paying every year, i.e. forever. These are MT solutions that indeed need re-training for some time, growing from childhood to adulthood but leaving home fairly soon.

- You will pay for first engine development and a few re-trainings, but the domain-/client-specific solution will be **yours** – and **fully customised!**

Thank you !

QUESTIONS ?

eyuste@pangea.com.MT

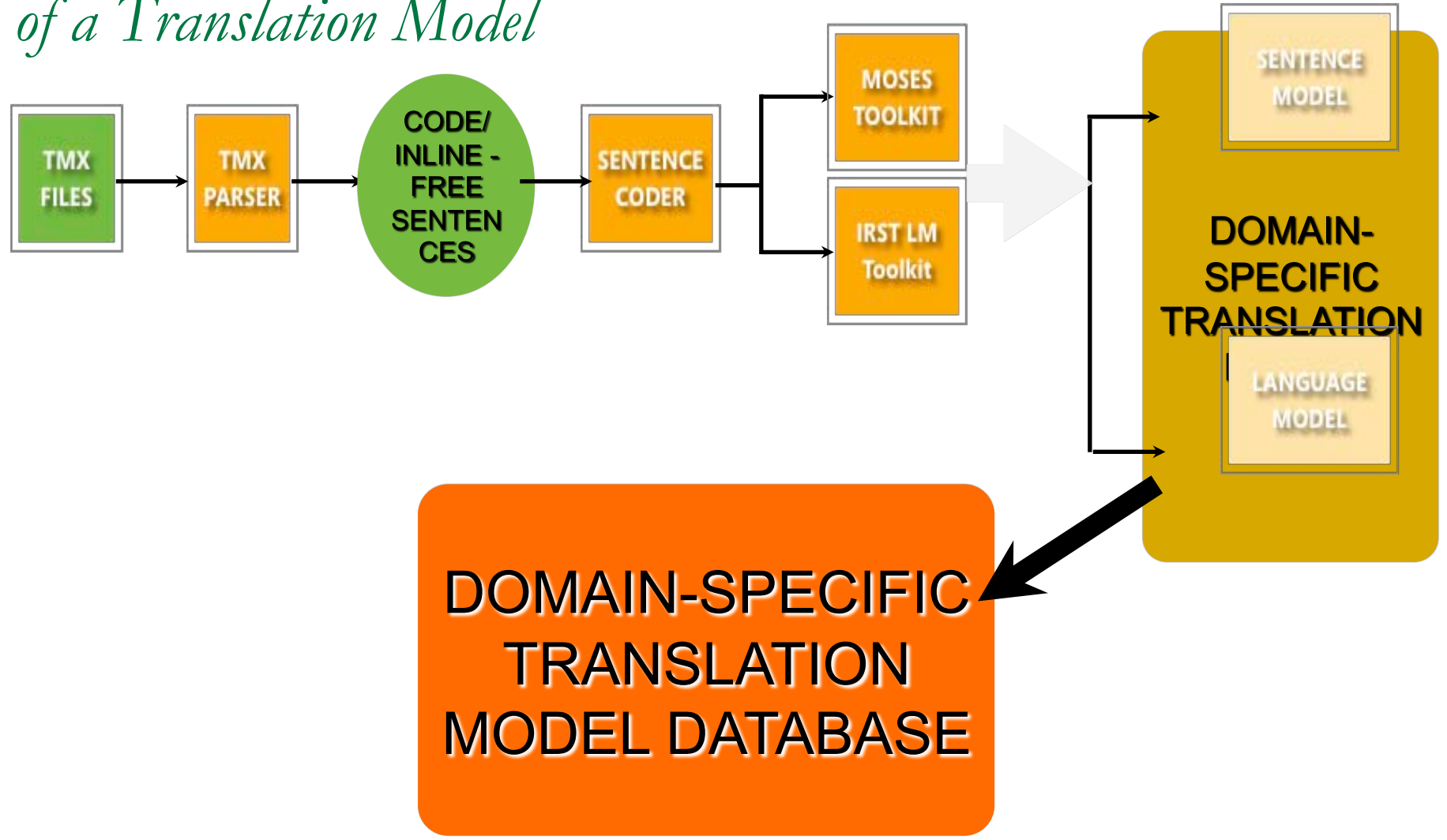
Your benefit

Drop us a line for a
complimentary 20-min phone consultation
to be scheduled within Q4 of 2010.

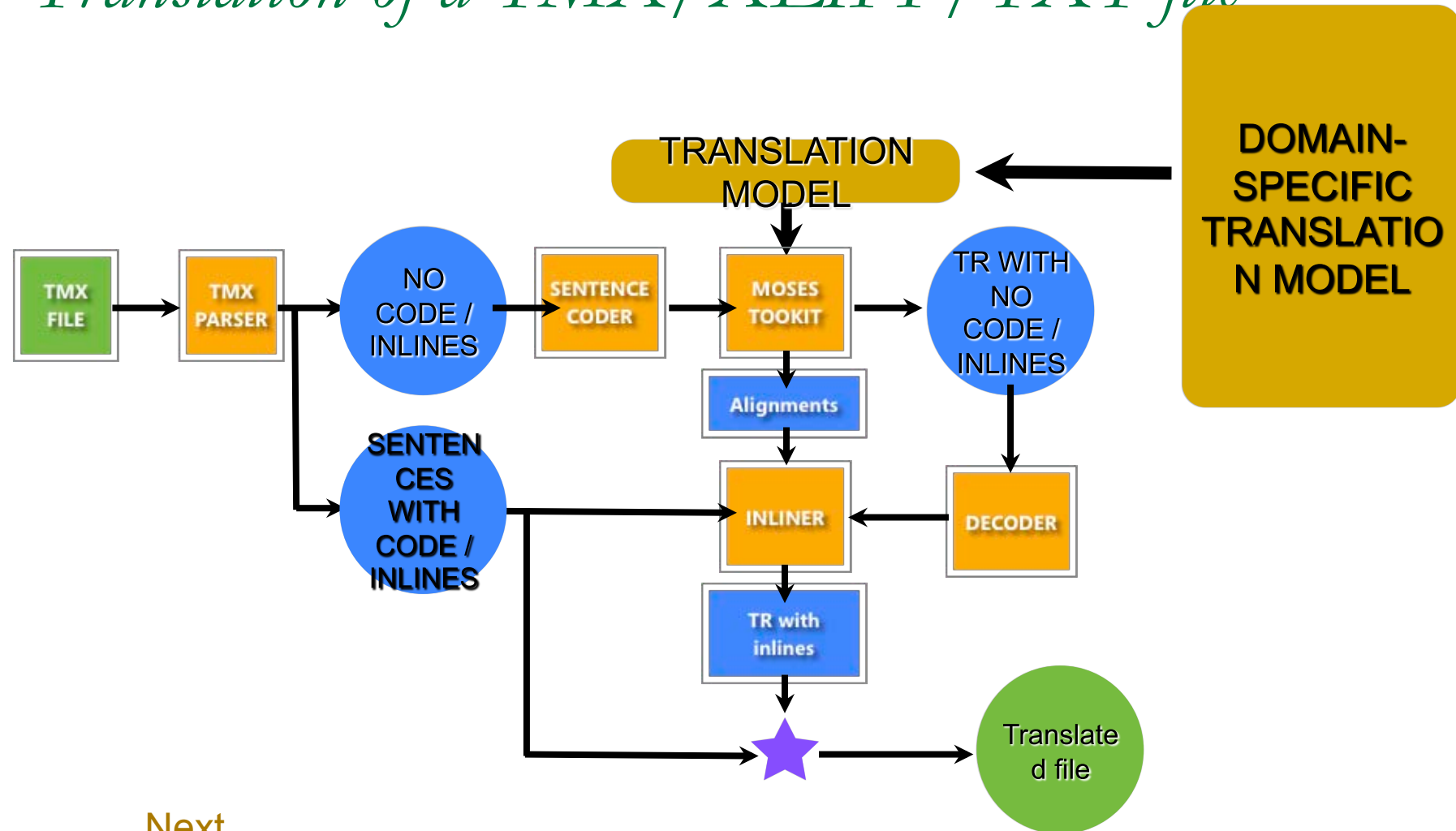
Please write **[After ATC'10 Conf]**
in your subject line.

Many thanks for your interest!

Appendix: What lies underneath (I) – *Training Process of a Translation Model*



Appendix: What lies underneath (II) – *Translation of a TMX/XLIFF/TXT file*



Next