

# PROMT at PayPal: enterprise-scale MT deployment for financial industry content

**Olga Beregovaya**

PROMT

330 Townsend Street, Suite 117

San Francisco, CA 94107

olga.beregovaya@PROMT.com

**Alex Yanishevsky**

PROMT

330 Townsend Street, Suite 117

San Francisco, CA 94107

alex@PROMT.com

## Abstract

This paper describes PROMT system deployment at PayPal including:

- PayPal localization process challenges and requirements to a machine translation solution
- Technical specifications of PROMT Translation Server Developer Edition
- Linguistic customization performed by PROMT team for PayPal
- Engineering Customization performed by PROMT team for PayPal
- Additional customized development performed by PROMT team on behalf of PayPal
- PROMT engine and PayPal productivity gains and cost savings.

## 1 Introduction

As a company, PayPal has always been at the top of the localization maturity model (LMM) due to many factors including: a tailored market approach for each region, a sophisticated workflow in SDL WorldServer and highly experienced personnel. PayPal's Localization team considered deployment

of machine translation as the next logical evolution of their localization process.

The following factors were PayPal requirements in deploying machine translation technology:

- High quality machine translation and proven ROI to help serve their localized markets cheaper and faster than ever
- Interaction with SDL Idiom WorldServer
- Flexibility in handling metadata
- On-the-fly locale conversion
- Independent engine management by client linguists and engineers
- Support for content types such as UI, On-line help, FAQ, User Guide, Error Messages, Marketing, Field offices feedback, User feedback
- Support for languages and locales such as English (US and UK) into/from French (Canada, France), Italian, German, Spanish (Spain, neutral), Portuguese (Brazil, Portugal), Russian, Simplified Chinese, Traditional Chinese

## 2 Addressing Client Challenges

Client challenges were addressed by PROMT's flagship enterprise product, PROMT Translation Server Developer Edition, as well as linguistic and engineering customization performed by the PROMT team for PayPal.

## 2.1 PROMT PTS Translation Server Developer Edition

PROMT's enterprise product's key capabilities are:

- Capacity for load distribution using multiple translation servers
- Advanced engine customization modules
- Client-server architecture, where all translations are performed on the server side
- All translation settings are managed on the client side and stored on the server side
- Administrative capabilities to control and enforce translation settings
- Secure data transmission
- Set of advanced pre- and post-processing algorithms
- Integration with GMS systems (SDL Idiom WorldServer in case of PayPal)
- Open and easily consumable API

## 2.2 Linguistic Customization

Linguistic customization included:

- identification and repair of inconsistencies between translation memory and glossary
- conversion of existing product glossaries into PROMT custom dictionaries
- extraction of new glossary candidates based on statistical harvesting methods using PROMT Terminology Manager and Content Maximization Suite
- coding of client glossary terms in PROMT Dictionary Editor
- Virtual Style Guide - upholding the client's stylistic guidelines.

The Virtual Style Guide is a compendium of syntactic choices. Depending on the context, the machine translation engine can resolve part of speech homonymy, amongst many other types of stylistic options, to yield the correct translation.

Example 1. Priority for the Imperative:

Source: *Log in to PayPal to add this email address to your account: has sent you a Money Request.*

Default: **Entrada** al sistema a PayPal para añadir esta dirección de correo electrónico a su cuenta: le ha enviado una Solicitud de dinero.

Imperative: **Entre a** PayPal para añadir esta dirección de correo electrónico a su cuenta: le ha enviado una Solicitud de dinero.

Example 2. Analysis of sentences with omitted subject:

Source: *Failed to update business name, the new name you have entered is blank or incorrect in some way.*

Default: **Fallado para** actualizar el nombre de la empresa, el nombre nuevo en el cual ha entrado es en blanco o incorrecto de algún modo.

Omitted subject: **Falló de** actualizar el nombre de la empresa, el nombre nuevo en el cual ha entrado es en blanco o incorrecto de algún modo.

Example 3. Translation for -ing-forms in temporal meaning:

Source: *All footnotes are positioned at bottom of page when converting from Word 2.0 to RFT-DCA.*

Gerund: *Todas las notas a pie de página están colocadas en el pie de página **convirtiendo** de Word 2.0 a RFT-DCA.*

Al+infinitive: *Todas las notas a pie de página están colocadas en el pie de página **al convertir** de Word 2.0 a RFT-DCA.*

Example 4. Translation of Passive

Source: *This email address has already been added to a PayPal account.*

Default: *Esta dirección de correo electrónico ha sido añadida ya a una cuenta de PayPal.*

Passive: *Esta dirección de correo electrónico se ha añadido ya a una cuenta de PayPal.*

Example 5. Swapping subject and predicate at the end of sentence.

Source: *You cannot capture funds because access to your account has been limited.*

Swap: *Usted no puede captar fondos porque ha sido limitado el acceso a su cuenta.*

Do not swap: *Usted no puede captar fondos porque el acceso a su cuenta ha sido limitado.*

## 2.3 Engineering Customization

The engineering customization performed by PROMT Professional Services on behalf of PayPal consisted of:

- Advanced metadata handling using PROMT-Idiom XLIFF Connector
- Handling variables using PROMT-Idiom XLIFF Connector, XML Rules, Dictionary Editor
- Pre- and post-processing rules

Plain text connectors strip out placeholders and metadata information. The tags are then randomly placed at the beginning and end of the sentence. This method results in significant detective work for the MT post-editors who now have to deal with matching up both tags and content.

Plain Text  
 Source: *To become verified and lift your sending limit, please confirm your email address, then add a credit or prepaid card to your PayPal account and {30} {31} {32} {33} {34} {35}confirm{36} {37} {38}it.{39}.*

Target: *{30}Para hacerse verificado y levantar su límite de envío, por favor confirme su dirección de correo electrónico, luego añada un crédito o tarjeta de prepago a su cuenta de PayPal y confírmelo.{31}{32}{33}{34}{35}{36}{37}{38}{39}:*

In contrast, in an XLIFF connector, all the placeholders are sent to the MT engine and are processed, while the semantic and syntactic integrity of the sentence remains intact.

Source: *To become verified and lift your sending limit, please confirm your email address, then add a credit or prepaid card to your PayPal account and {30} {31} {32} {33} {34} {35}confirm{36} {37} {38}it.{39}.*

Target: *Para hacerse verificado y levantar su límite de envío, por favor confirme su dirección de correo electrónico, luego añada un crédito o tarjeta de prepago a su cuenta de PayPal y {30} {31}{32}{33}{34}{35}confírmelo{36}{37}{38}. {39}*

Consequently, better inline parsing by the MT engine yields better, more natural MT output which in turn leads to reduced efforts for post-editors and more productivity and reduced costs.

Since PayPal's core business requires processing financial transactions, the source content contains a multitude of shortcuts or variables that are populated at runtime. It is not uncommon to see sen-

tences such as:

<1> <2> paid <3> from his/her <4> <5> to <6> <7> on <8>.

John Smith paid \$100 from his/her German bank account to Jane Doe on January 1, 2010.

Such challenges are addressed by the flexibility of PROMT's engine. The XML Rules module provides full control in demarcating the tags as "translate" or "do not translate." The Dictionary Editor module controls the handling of variables; in the example above, <4> is coded as a variable and as an adjective. Thus, the variable follows the language grammar and yields a correct translation: "compte bancaire allemande."

PROMT introduced a set of pre- and post-processing algorithms to handle symbols such as ', & and " by doing substitutions so the symbols conform to the XML standard. In addition, the grammatical parsing/reordering of some languages creates more tags than the original and Idiom will not accept this solution.

For example, the following translation from English into German:

*It <b>has worked</b> very well. (two tags, one open and one close bold)*

*Es <b>hat</b> sehr gut <b>gearbeitet</b>. (four tags, two open and two close bold).*

The XLIFF connector verifies that the number of placeholders in the translated segment is equal to the number of placeholders in the source segment; if they are not equal, the algorithm strips out the extraneous tags.

## 2.4 On-the-fly locale conversion – PROMT Normalization Suite

Single sourcing or repurposing content has become a mainstay in today's highly competitive environment. At the same time, PayPal needs to differentiate or localize its content for each region. Out of these two contradictory requirements, a truly unique localization solution, the Normalization Suite, was born. PROMT created a suite of linguistic tools that allow its machine translation engine to convert on-the-fly between same languages with different locales such as English (US) <> English (UK), French (France) <> French (Canada), Spanish (Spain) <> Spanish (neutral).

Example of US English to UK English real-time conversion:

*Check that your **check** was sent in the US mail.  
Check that your **cheque** was sent in the US post.*

## 2.5 Ongoing linguistic and engineering maintenance by PayPal

Ongoing maintenance of PROMT engine is achieved through a combination of the following:

- white box philosophy to the product
- an intuitive UI
- adherence to standards that allows for ease of customization and ongoing maintenance
- in-depth training of PayPal Localization team members

PayPal Localization team has full access to engine customization modules and is able to make real-time changes independently, only engaging PROMT Professional Services for additional development and advanced linguistic coding.

## 3 Next Steps

Currently, PROMT is working on deploying two additional core features: PROMTrust and a Deep Hybrid module; both of these new features will be deployed at PayPal.

PROMTrust Provides a reliability score/confidence factor based on user dictionary and TM markup. It is useful for files where post-editing will be done since higher scores correlate with reduced post-editing.

PROMT Deep Hybrid approach increases the relevance of the final translation output and reduces the post-editing effort since it implements a combination of rules-based and statistical methods at the very core of translation process, not just for post-editing the final output. The module uses the client's linguistic assets at all stages of the translation process and chooses a candidate based on the lowest level of perplexity.

## Example 1: Syntactic choice

### Source:

It **is used** for patient information, lab results, reports, images, and clinical data.

### RBMT translation:

**Es usado** para información sobre los pacientes, resultados del laboratorio, informes, imágenes, y datos clínicos.

### Hybrid engine candidates:

a) **Es usado** para información sobre los pacientes, resultados del laboratorio, informes, imágenes, y datos clínicos.

*ppl= 791.4319204909*

b) **Se usa** para información sobre los pacientes, resultados del laboratorio, informes, imágenes, y datos clínicos.

*ppl= 424.83820234214*

c) **Está usado** para información sobre los pacientes, resultados del laboratorio, informes, imágenes, y datos clínicos.

*ppl= 814.24328845084*

### Hybrid Outcome:

**Se usa** para información sobre los pacientes, resultados del laboratorio, informes, imágenes, y datos clínicos.

## Example 2: Lexical choice

### Source:

The "Nehalem" system architecture **features** an integrated memory controller

### RBMT translation:

La arquitectura del sistema "Nehalem" **presenta** un controlador de memoria integrado

### Hybrid engine candidates:

a) La arquitectura del sistema "Nehalem" **presenta** un controlador de memoria integrado

*ppl= 288.17916810444*

b) La arquitectura del sistema "Nehalem" **incluye** un controlador de memoria integrado

*ppl= 234.86938828311*

### Hybrid Outcome:

La arquitectura del sistema "Nehalem" **incluye** un controlador de memoria integrado

## 4 Conclusion

As a result of the linguistic and engineering customization described above, PayPal was able to increase throughput productivity by approximately 30%.