

**Booking.com**

**Maxim Khalilov**

# **Evaluation of machine translation quality in e-commerce environment**

**AMTA 2016**  
**Commercial MT Users and Translators**

**October 31, 2016**  
**Austin, TX**

## Booking.com story.

The world's #1 website for booking hotels and other accommodations.

- Founded in 1996 in Amsterdam
- Part of the Priceline Group (NASDAQ: PCLN) since 2005
- 1,000,000+ properties in more than 200 countries including 490,000+ vacation rental properties
- Over 1,100,000 room nights every 24 hours
- Accommodation available in: 220+ countries and territories
- Number of unique destinations worldwide: 94,000+
- 42 languages
- 184 offices worldwide in 50+ countries
- More than 13,000 employees



Booking.com

## Outline.

- Why? MT for e-commerce.
- MT at Booking.com: timeline and approach
- Why MT can be dangerous and how can we deal with it?
- Evaluation of MT quality: methods
- Quality assurance system
- Benefits
- Q&A



## Market strategy

Growth is driven by the localization of global market strategies



## Technology

Significant impact on the e-commerce business in general by enabling global customers reach the product description content across language borders

## Industry

Travel industry was one of the first to come to the Web



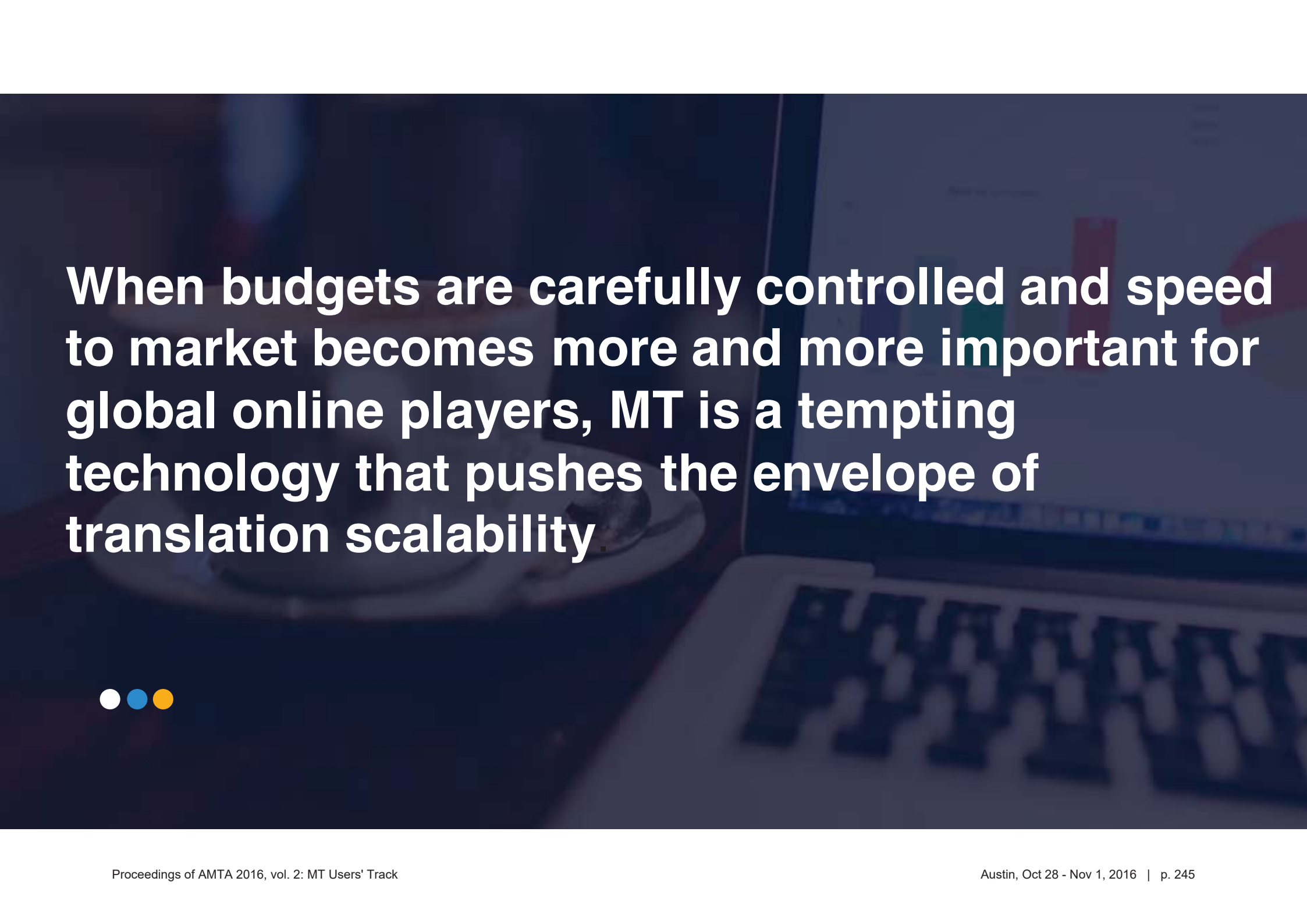
## Languages

Demand-driven language selection



# Why Machine Translation?





**When budgets are carefully controlled and speed to market becomes more and more important for global online players, MT is a tempting technology that pushes the envelope of translation scalability.**



## Four business cases. (for Booking.com)

### 1. Property descriptions

MT helps to scale property (hotels, apartments, hostels, etc.) descriptions into 42 languages.

### 2. Translation support

In-house MT supports translation department of the company to increase the productivity of translation process.

### 3. Customer experience applications

Our guests and partners consumer and produce content in different languages in various forms from USGs to CS tickets.

### 4. Other cases

Various other MT use cases, including Big Data applications and user feedback processing.

**Booking.com**

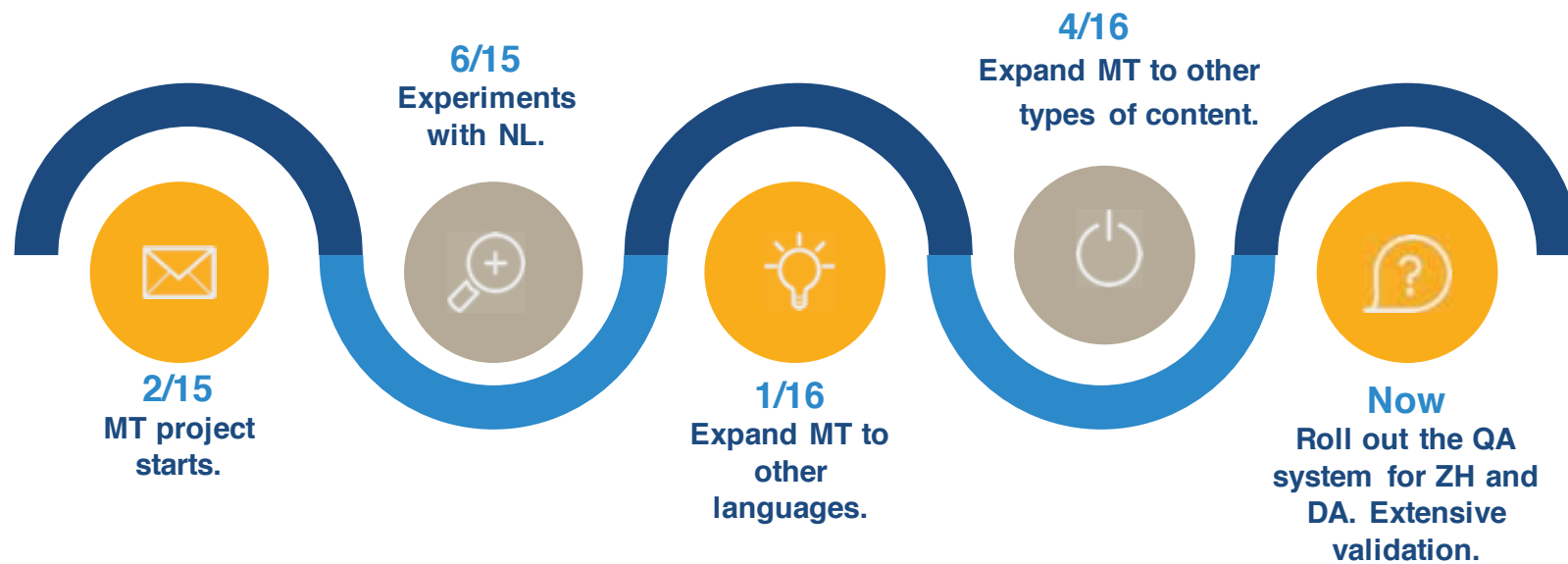
Property  
descriptions.

Translation  
support.

Customer  
experience  
applications.

Other cases.

# MT timeline.



Booking.com





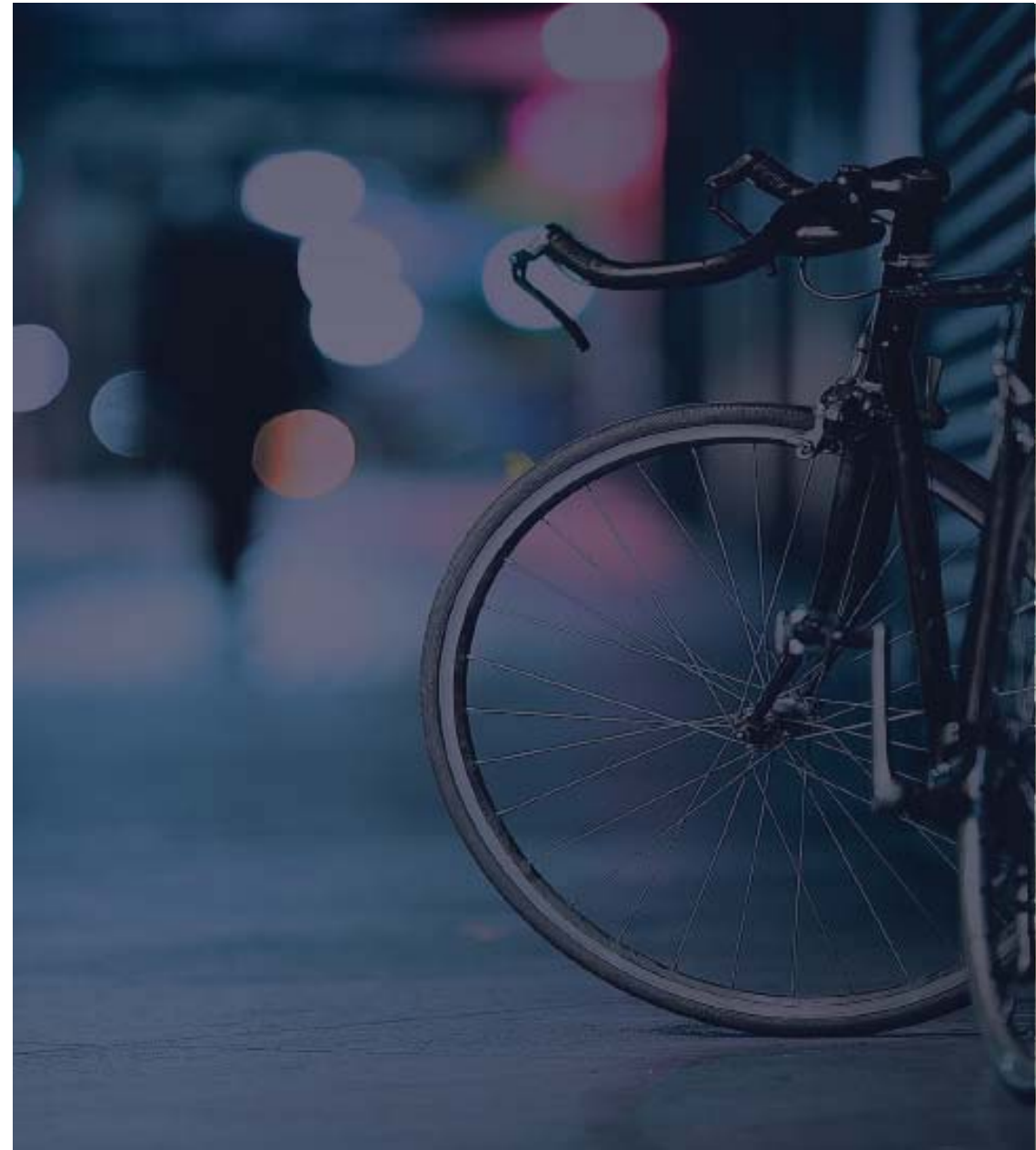
## Why MT can be dangerous?

**The imperfection of MT might mislead users, have legal consequences for the company or damage brand's reputation and customer's confidence of translated content.**

**Booking.com**

## Challenges

- A lack of objective and comprehensive methodology of MT quality evaluation that would be flexible enough to give reliable results for different types of content.
- An absence of the clear link between a fully automatic metric or a set of metrics with business impact of translation quality.



# How can we control MT quality in e-commerce environment?



# Human evaluation.



Error analysis

Useful for regular checks of translation quality from the linguistic perspective



Usability assessment  
Adequacy/Fluency scoring

Rough assessment of the MT-ed content in terms of its publishability



User feedback  
User behavior analysis

Both methods are mostly used in the customer-facing scenarios



Productivity measurement

Post-editing MT scenario

Booking.com

# Automatic methods.



Conventional  
automatic metrics



Applicable to make sure  
there are no new bugs  
introduced as the result  
of the MT engine  
retraining.



Number of business-sensitive  
errors



Methods that link MT  
quality with potential  
threats for the business

Booking.com



## Problem.

- **Automatic metrics for MT quality evaluation, such as BLEU, METEOR and GTM are not capable of reflecting the usability of MT-ed content**
- **They also can not distinguishing harmful errors from those that do not have critical impact on the overall translation understandability and adequacy.**
- **Human quality assurance is time consuming and non-cost effective.**
- **Human evaluation is subjective.**

**Booking.com**

# Quality assurance system



# How?





## Solution.

**Quality heuristics: enhance the quality evaluation model to ensure the MT system is not making sensitive errors like offering free facilities that aren't actually free, or mistranslating distances.**

# Implementation.

## Sensitive words.

- 
- 
- 
- 

Manually created multilingual list of sensitive words, expressions and values that can prejudice the accuracy of translation

## Matching module.

- 
- 
- 
- 
- 
- 
- 
- 

Compares factual and numerical discrepancies between source copy and MT output

## Reaction.

- 
- 
- 
- 
- 
- 

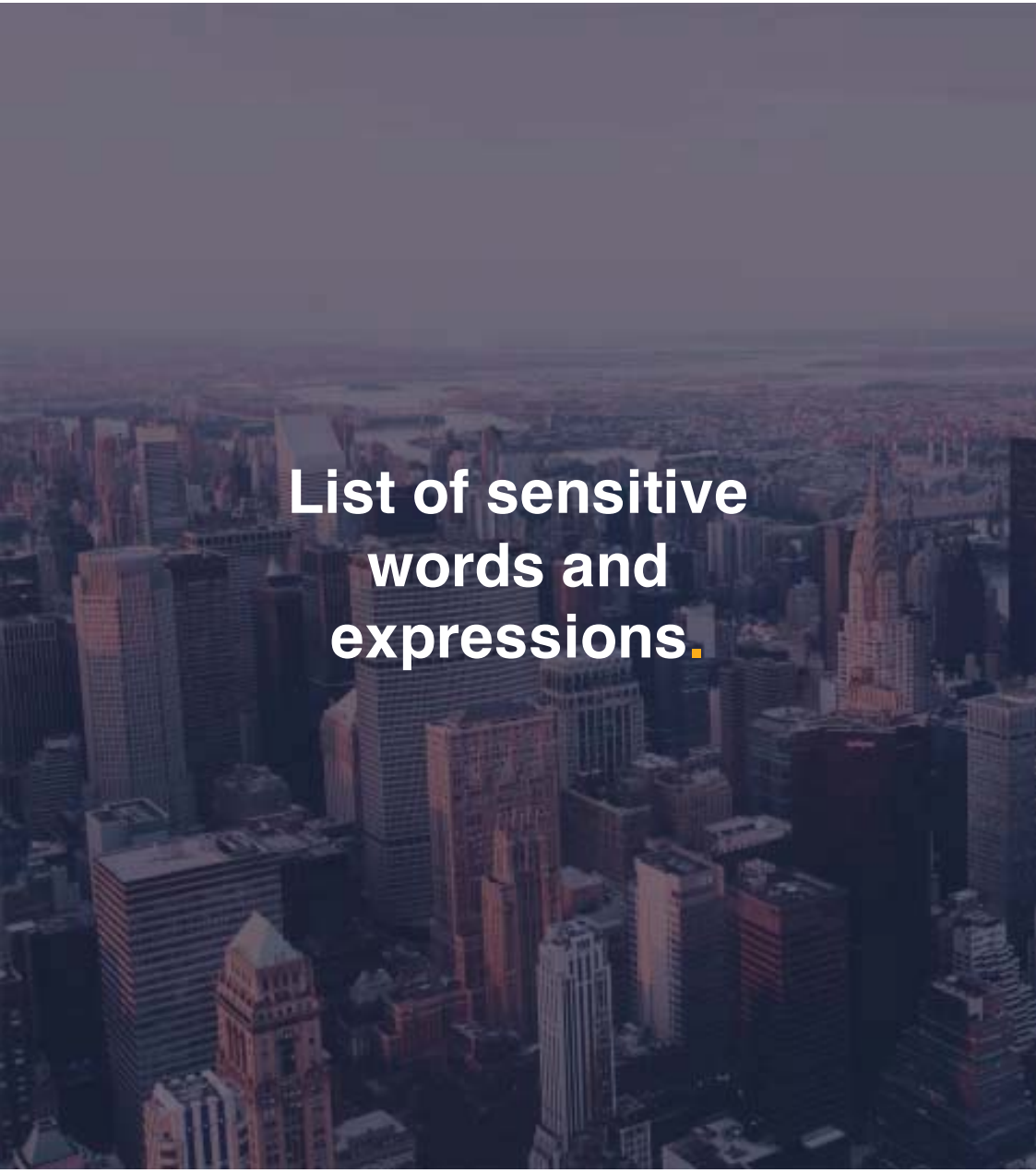
Actions in return to a “defective” translation

## Feedback loop.

- 
- 
- 
- 

Process of multilingual list correction and modification

Booking.com



## List of sensitive words and expressions.

### Several categories:

- Free/non-free
- Available/unavailable
- Payment and price
- Location and distance
- Time
- Units of measure
- Other

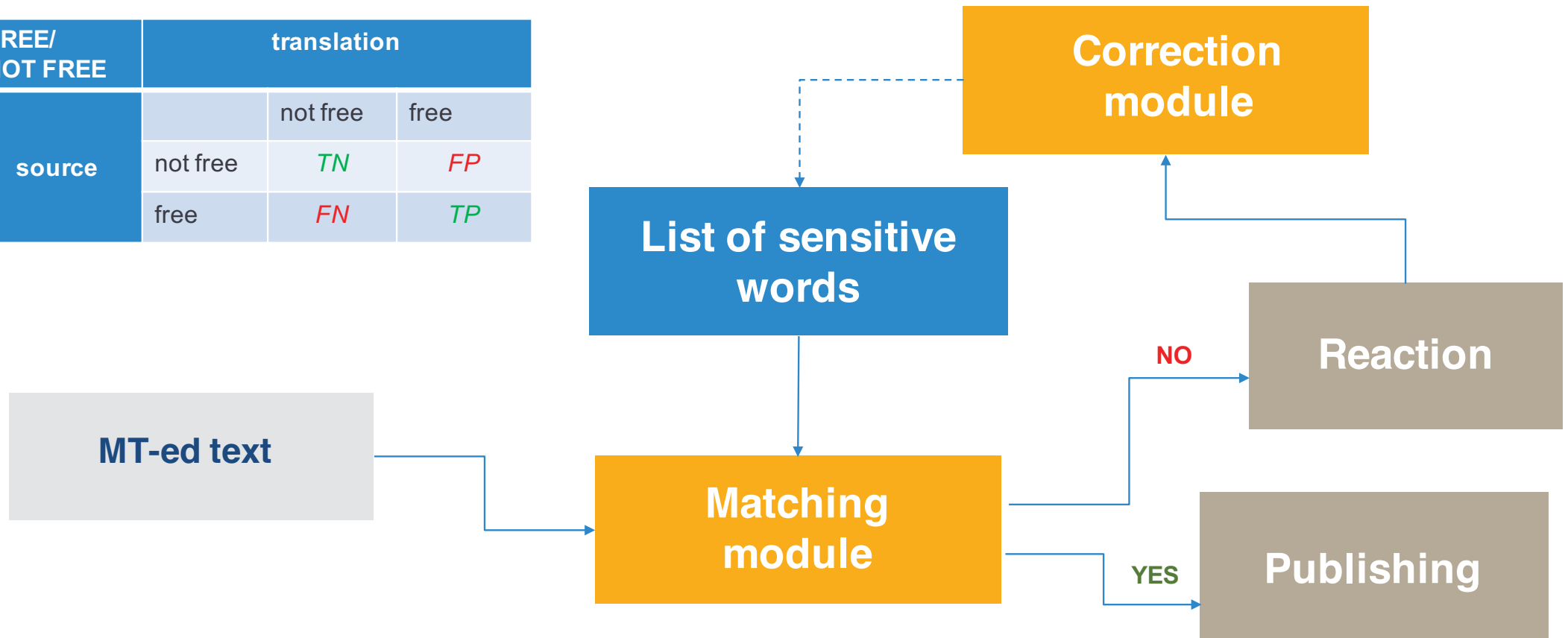
### Statistical testing:

- False positives and false negatives

[Booking.com](https://www.booking.com)

# Quality assurance system

FREE/ NOT FREE	translation		
	not free	free	
source	not free	<i>TN</i>	<i>FP</i>
	free	<i>FN</i>	<i>TP</i>



# Advantages?





## Benefits.

1. **Measure the quality of MT for web publishing in a meaningful way**
2. **Minimize subjectivity of MT quality evaluation**
3. **Ensure the accuracy of MT-ed content (with a certain level of confidence)**
4. **Gives a quantifiable measure of business impact caused by some of MT failures**
5. **Increases translators engagement**

**Booking.com**



# Thank you!

**Booking.com**

All references to "Booking.com", including any mention of "us", "we" and "our" refer to Booking.com BV, the company behind Booking.com™