# Multidimensional Quality Metrics: A Flexible System for Assessing Translation Quality

Aljoscha Burchardt, Arle Lommel & Hans Uszkoreit (DFKI)

Funded by the 7th Framework Programme of the European Commission through the contract 296347.

QT LAUNCH PAD

# PROBLEMS IN ASSESSING QUALITY

QT LAUNCH PAD

95% of professionally translated content at one major

LSP is never evaluated for quality.

QT LAUNCH PAD

"Translators are the garbage collectors of the documentation world" –Alison Toon, HP

LAUNCH PAD

# “I know it when I see it”

QT LAUNCH PAD

Machine Translation quality scores (BLEU, NIST, etc.) are totally different from human evaluation.

QT LAUNCH PAD

MT methods require

reference translations:

Cannot be used for

production purposes

QT LAUNCH PAD

# Change the reference translation(s) and the score changes

QT LAUNCH PAD

# The problem with BLEU

No substantial improvement for human use

# The problem with BLEU

Substantial human improvement but no BLEU improvement

Human quality assessment

takes too much time

and is not principled

QT LAUNCH PAD

Most sampling today
is random but

errors are not random

Wait a minute...

# What do you mean by quality?

QT LAUNCH PAD

# Quality: A New Definition

A quality translation demonstrates
required accuracy and fluency
for the audience and purpose and
complies with all other negotiated specifications,
taking into account end-user needs.

Source: Alan Melby

# Sounds simple, right?

QT LAUNCH PAD

It's actually quite radical
and it drags translation
kicking and screaming
into the modern world of
quality management

QT LAUNCH PAD

# Multidimensional Quality Metrics

QT LAUNCH PAD

# Why not use a single metric everyone shares?

QT LAUNCH PAD

OK.

Which one?

LISA QA Model
SAE J2450
SDL TMS
Acrocheck
ApSIC XBench
CheckMate
QA Distiller
XLIFF:Doc
EN15038...

LAUNCH PAD

# All of them disagree* about what is important to quality

*The only thing they agree on is terminology

QT LAUNCH PAD

(Probably because there is

no single set of criteria

that applies to all kinds of

translation)

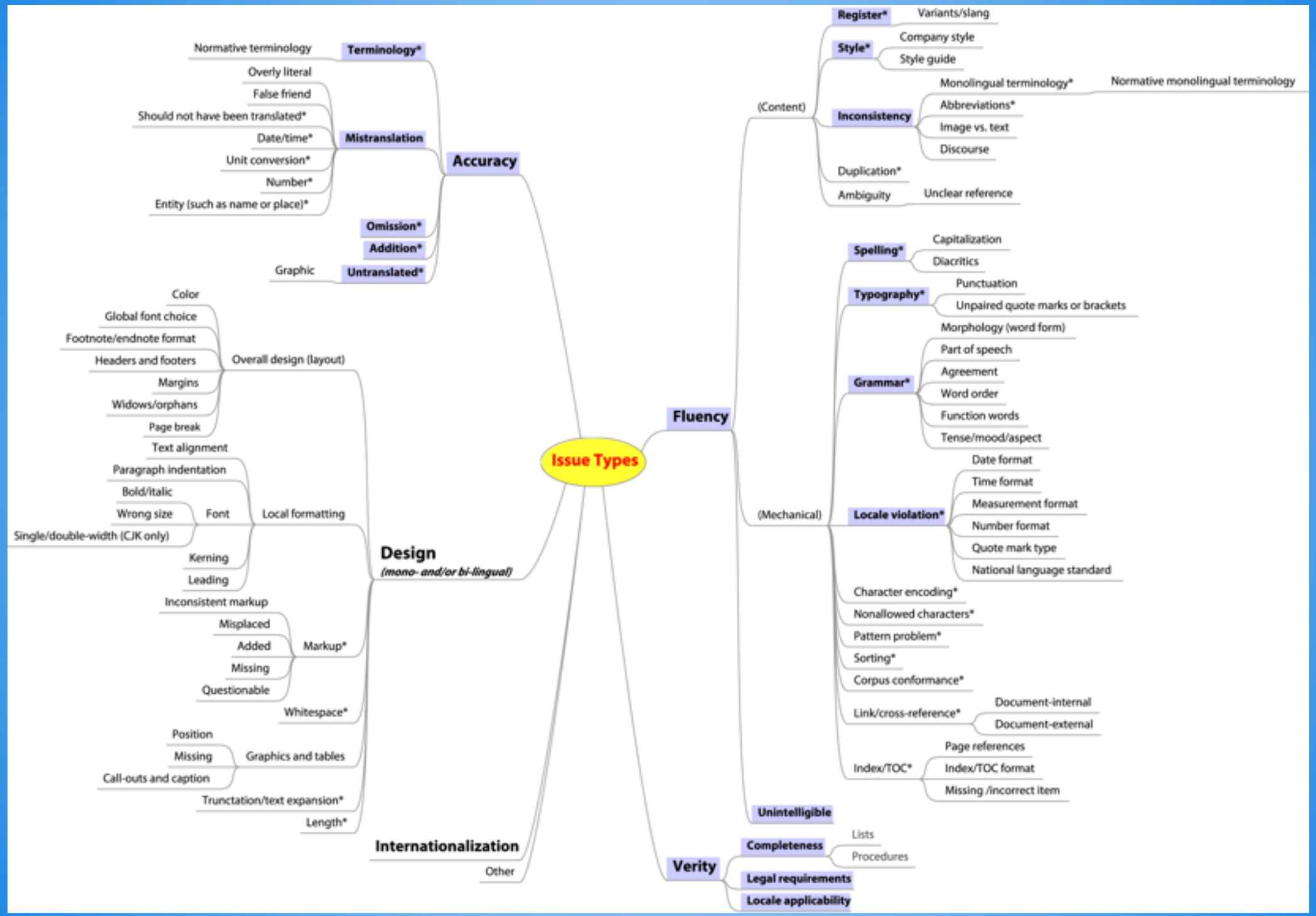
QT LAUNCH PAD

# There is no one-size-fits-all metric

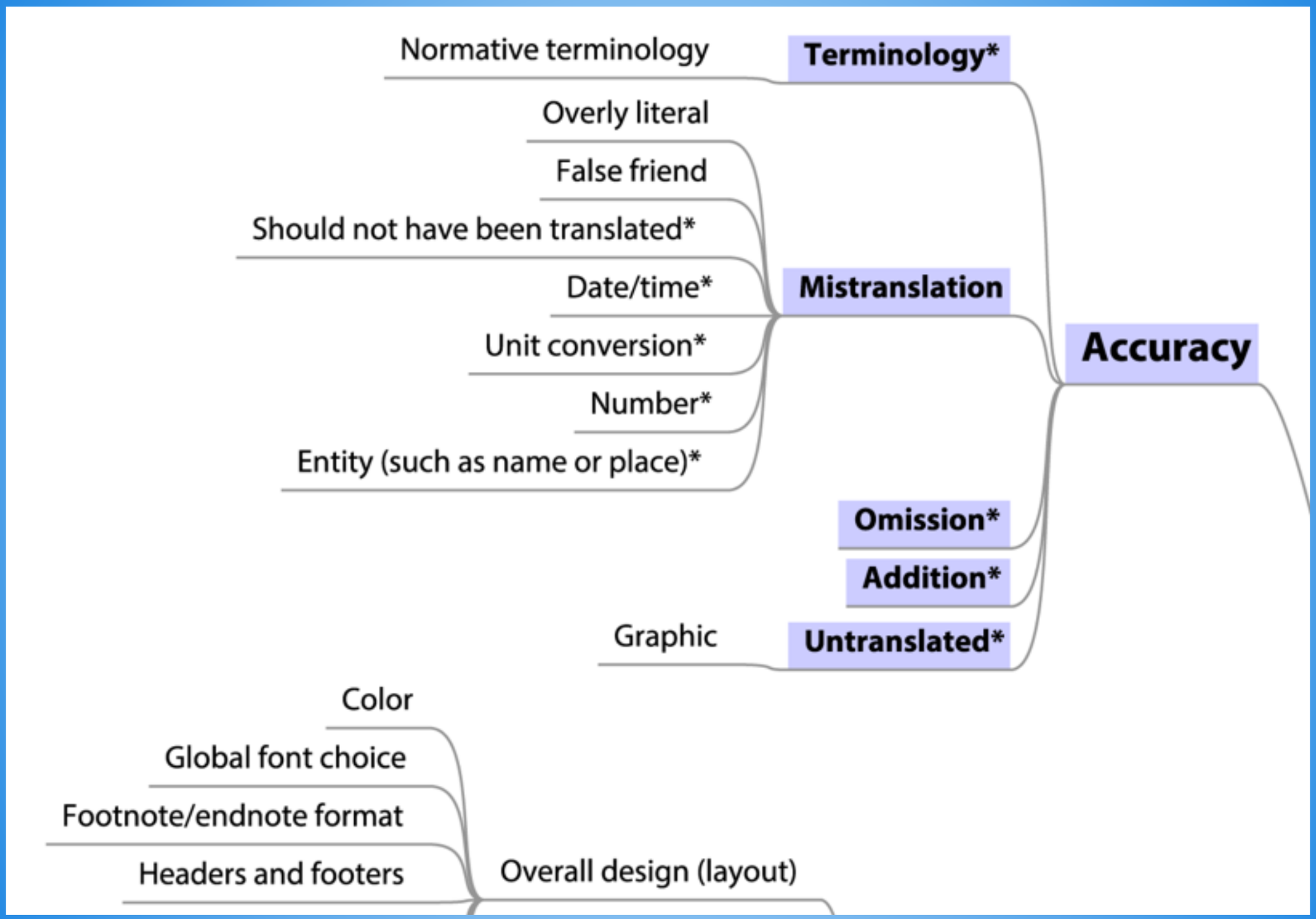
LAUNCH PAD

MQM provides a catalog of issue types suitable for various tasks

QT LAUNCH PAD



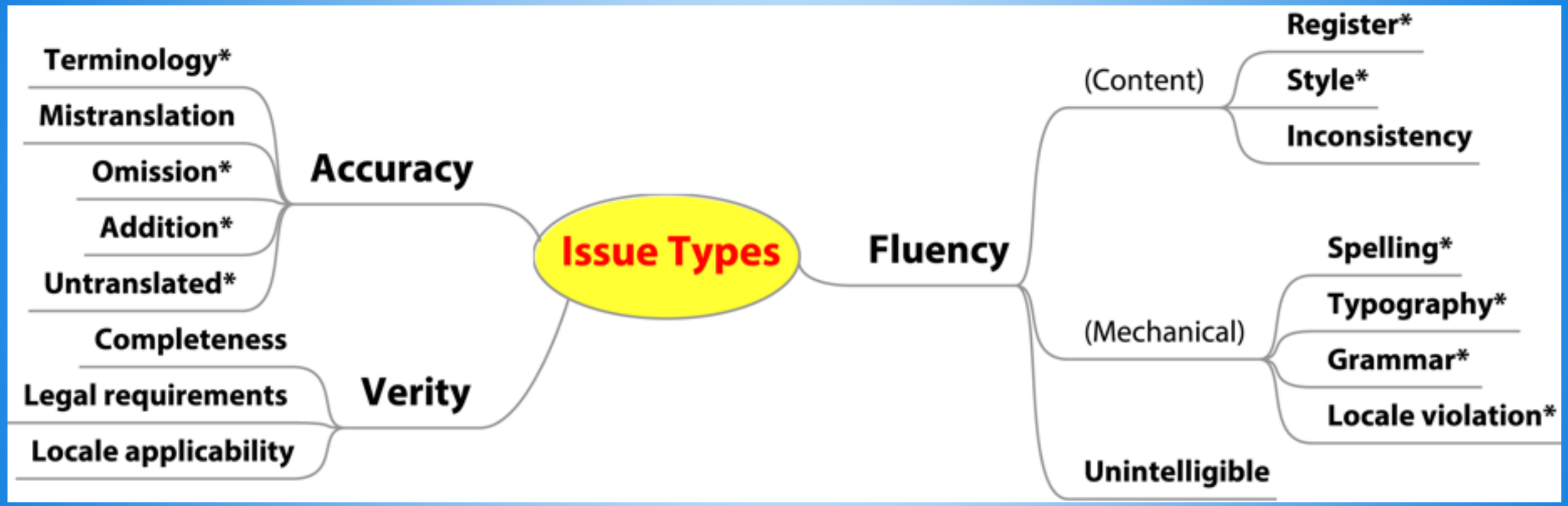

# The "full" MQM ("Spaghetti")




# Zooming into the "full" MQM

# Wait! Weren't we trying to improve things?

QT LAUNCH PAD


Issue Types
Accuracy
Terminology*
Mistranslation
Omission*
Addition*
Untranslated*
Completeness
Verity
Legal requirements
Locale applicability
Fluency
(Content)
Register*
Style*
Inconsistency
(Mechanical)
Spelling*
Typography*
Grammar*
Locale violation*
Unintelligible

# The MQM Core

Issues marked with a star (*)
can be automatically
detected

QT LAUNCH PAD

# Accuracy and Fluency

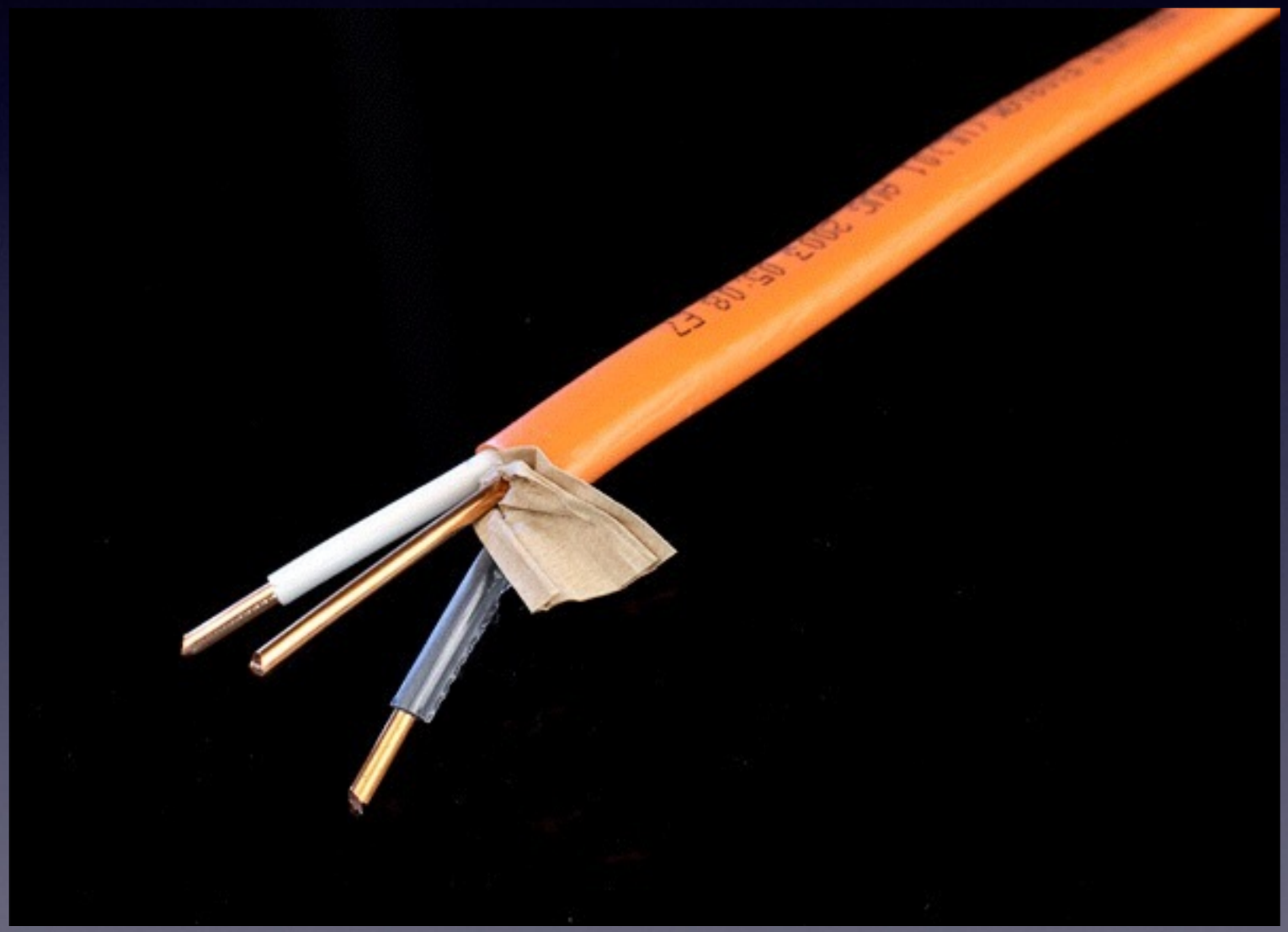
# What's Verity?

QT LAUNCH PAD

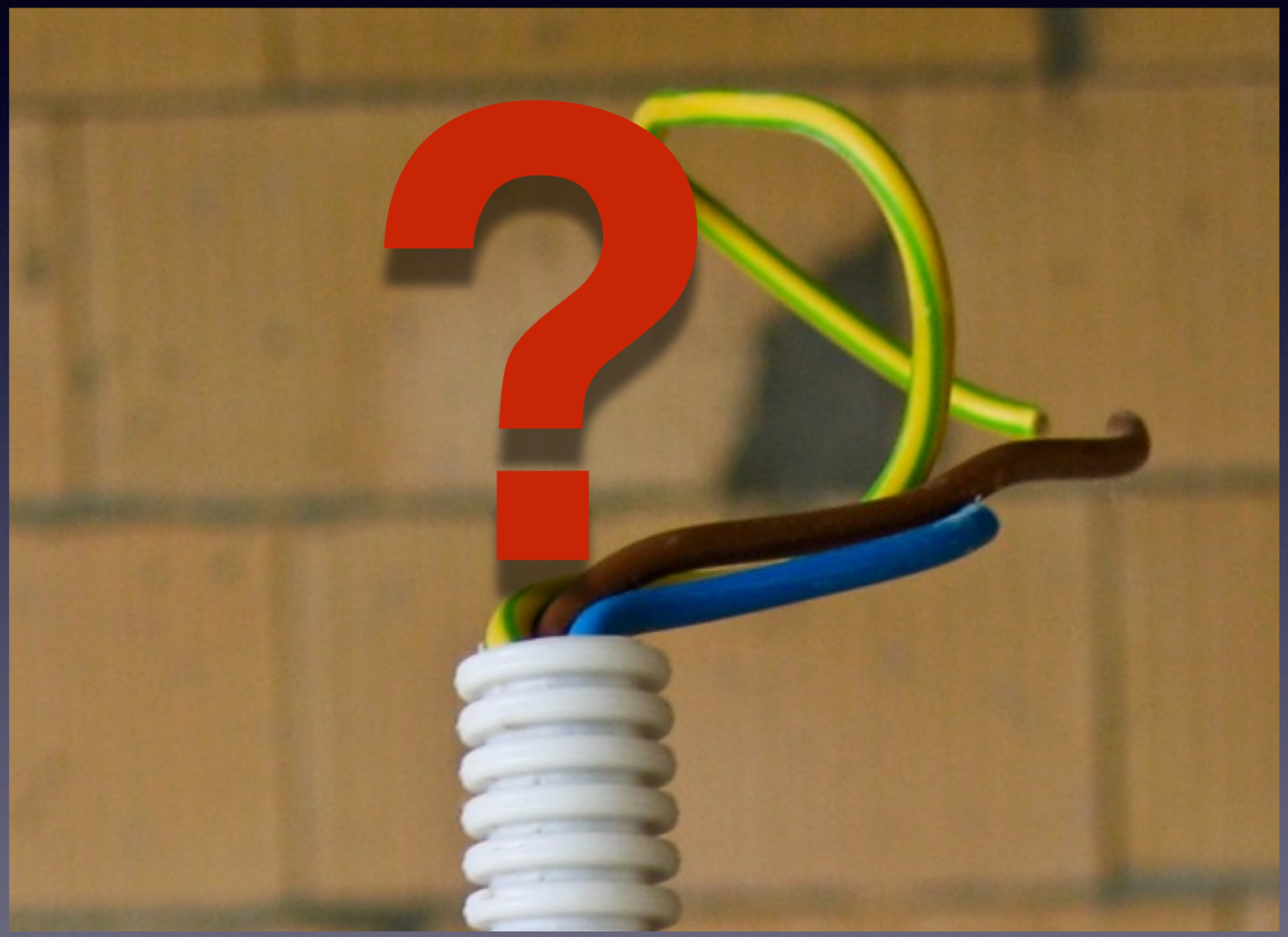
Verity provides a way to deal with the text in relation to the real world.

QT LAUNCH PAD

# Locate the ground wire (bare copper)...



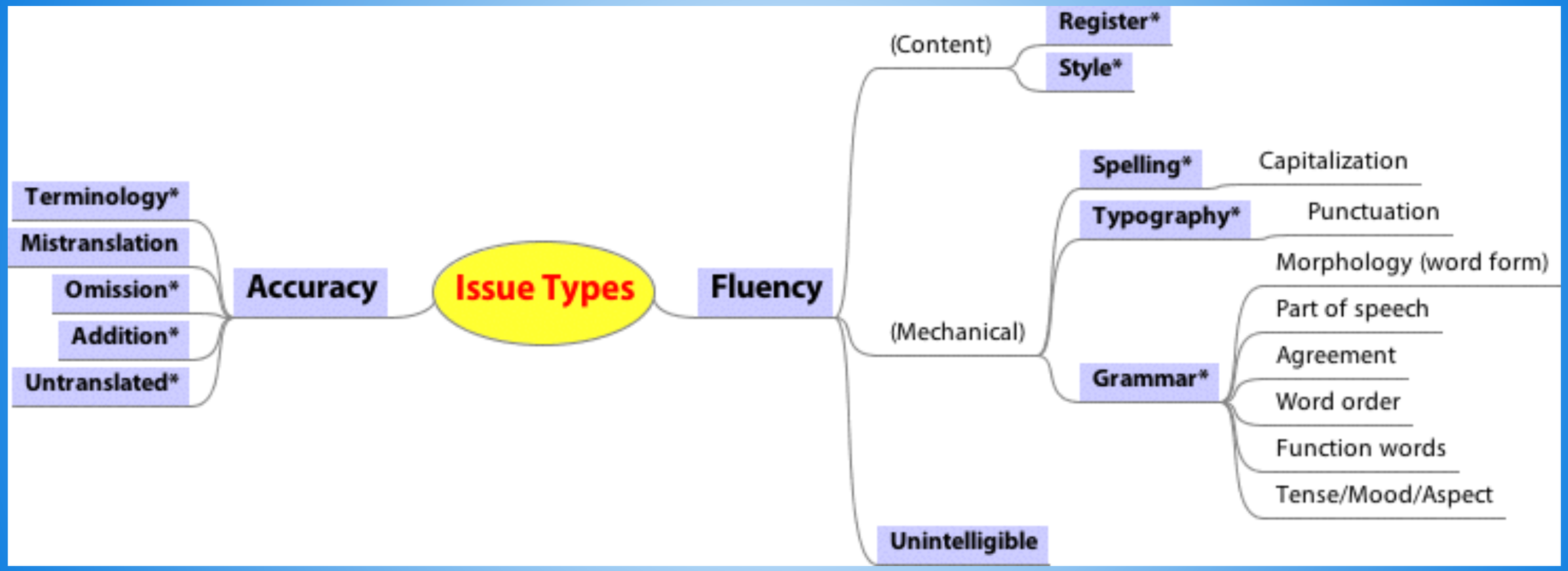
# Finden Sie das Erdungskabel (blankes Kupfer)



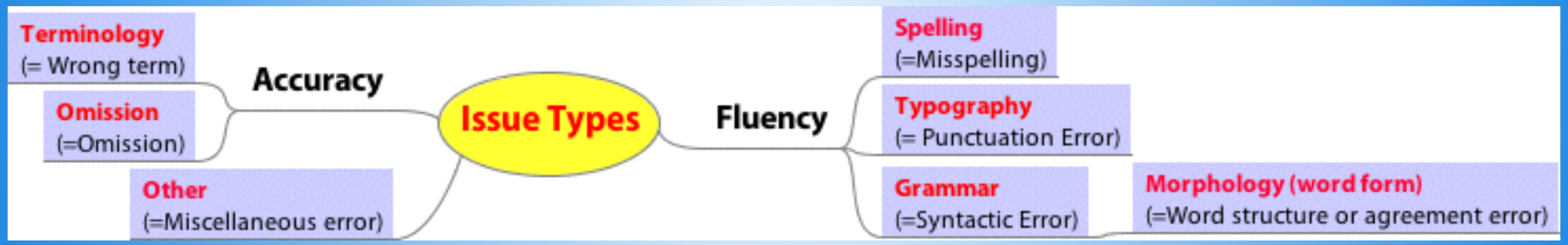
# You don't use all of MQM (or its core):

# you use the parts you need.

QT LAUNCH PAD




# MQM for MT Diagnostics


Issue Types
Accuracy
Terminology
(= Wrong term)
Omission
(=Omission)
Other
(=Miscellaneous error)
Fluency
Spelling
(=Misspelling)
Typography
(= Punctuation Error)
Grammar
(=Syntactic Error)
Morphology (word form)
(=Word structure or agreement error)

# SAE J2450

MQM lets you declare your quality metric in a shared vocabulary.

QT LAUNCH PAD

Dimensions help you decide what to check (and also help you communicate with your LSP)

QT LAUNCH PAD

No more assuming what the parties want or how to check it

QT LAUNCH PAD

# 12 Dimensions

(from ISO/TS-11669)

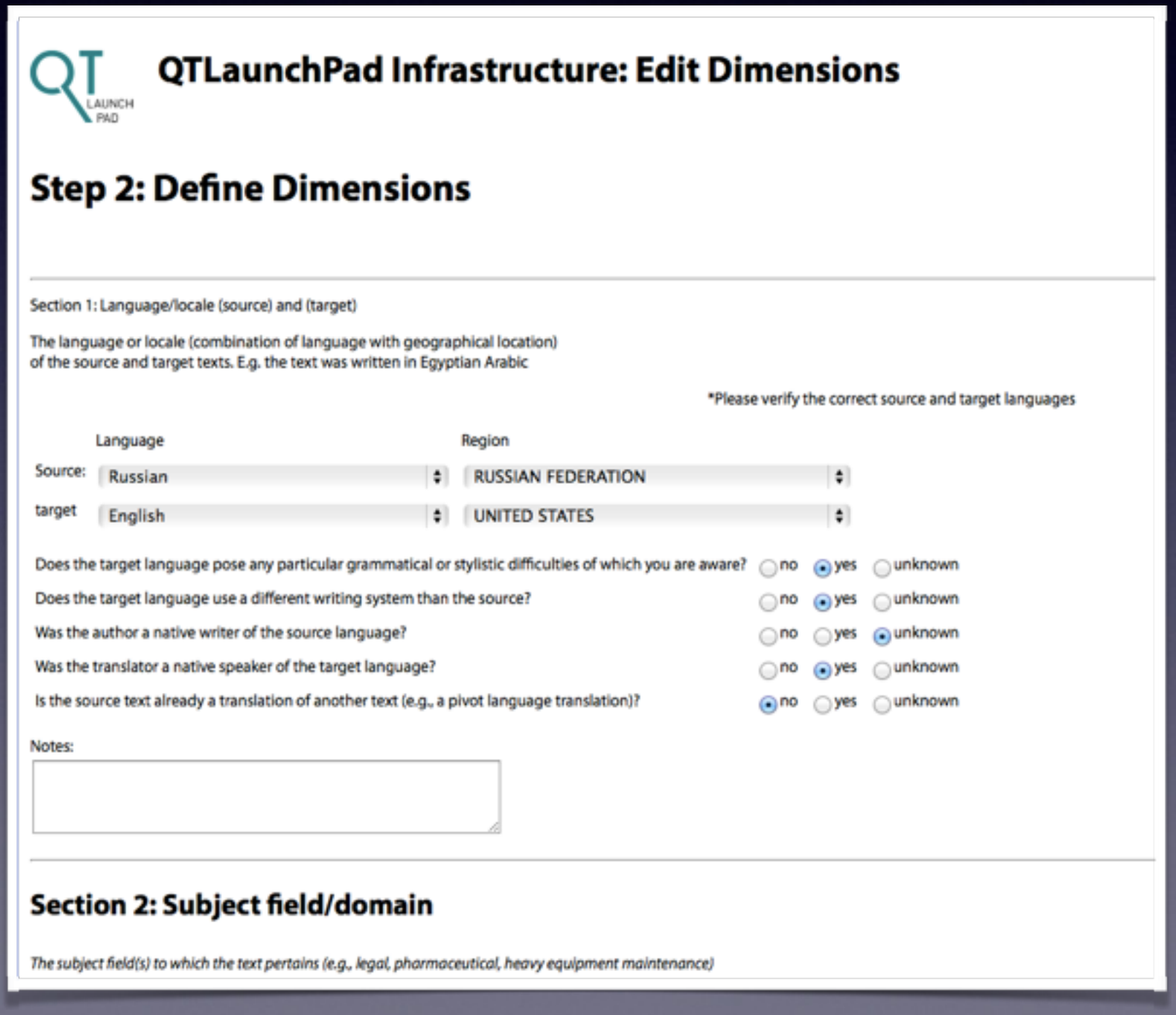
1. Language/locale
2. Subject field/domain
3. Terminology (source/target)
4. Text type
5. Audience
6. Purpose
7. Register
8. Target text style
9. Content correspondence
10. Output modality
11. File format
12. Production technology

# Open-source tools* to demonstrate MQM

*translate5 source code is published.
Other tools' code will be published in 2014

QT LAUNCH PAD

# An Online Tool for Building Dimensions and Metrics



QTLaunchPad Infrastructure: Edit Dimensions

## Step 2: Define Dimensions

Section 1: Language/locale (source) and (target)

The language or locale (combination of language with geographical location) of the source and target texts. E.g. the text was written in Egyptian Arabic

*Please verify the correct source and target languages

| | Language | Region |
|---|---|---|
| Source: | Russian | RUSSIAN FEDERATION |
| target | English | UNITED STATES |

| | | | |
|---|---|---|---|
| Does the target language pose any particular grammatical or stylistic difficulties of which you are aware? | no | yes | unknown |
| Does the target language use a different writing system than the source? | no | yes | unknown |
| Was the author a native writer of the source language? | no | yes | unknown |
| Was the translator a native speaker of the target language? | no | yes | unknown |
| Is the source text already a translation of another text (e.g., a pivot language translation)? | no | yes | unknown |

Notes:

## Section 2: Subject field/domain

*The subject field(s) to which the text pertains (e.g., legal, pharmaceutical, heavy equipment maintenance)*

http://www.qt21.eu/MQM

## Section 2: Subject field/domain

*The subject field(s) to which the text pertains (e.g., legal, pharmaceutical, heavy equipment maintenance)*

Subject Fields:

Is the subject field in a regulated industry where legal compliance is mandated? ○ no ◉ yes ○ unknown

Is the subject field a technical field where particular terminology is expected? ○ no ◉ yes ○ unknown

Notes:

---

## Section 3: Terminology (source/target)

*Any special terminology that needs to be noted in the source or target texts*

Terminology (list or reference):

volga.tbx

Notes:

---

## Section 4: Text type

*The type/genre of the content (e.g., novel, technical user manual, marketing)*

Text type:

service manual

Is the text type likely to require special attention to style or accuracy? ○ no ◉ yes ○ unknown

Will the text type have any other implications for other aspects of the translation? ◉ no ○ yes ○ unknown

Notes:

# Tabular Scorecard

Show Dimensions

**Score card**

The following score card automatically calculates scores based on the information entered.

| Information about the text | | Severity Multipliers | |
|---|---|---|---|
| Word Count (Source): | 200 | Minor: | 1 |
| Word Count (Target): | 300 | Major: | 5 |
| | | Critical: | 10 |

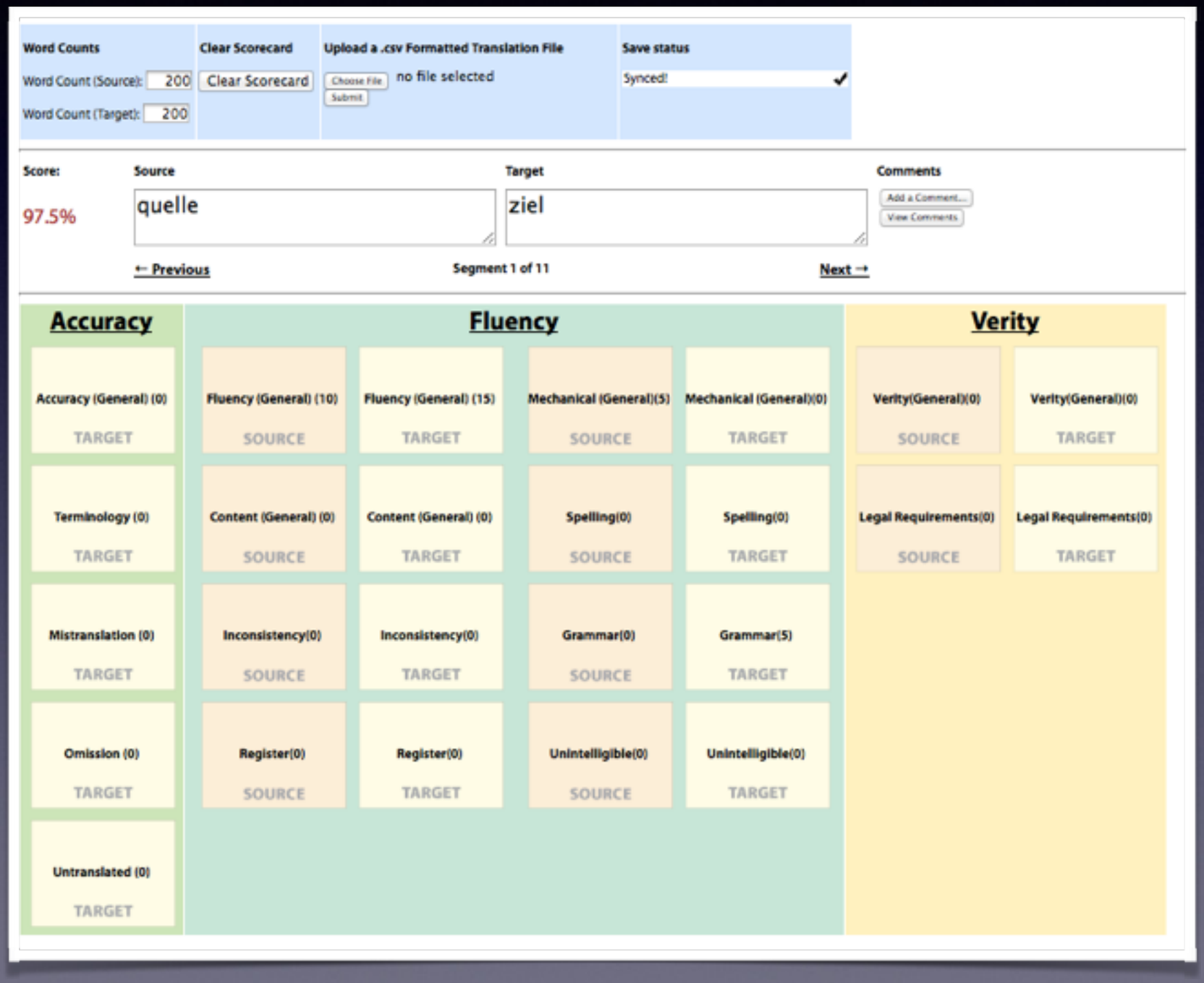
**95.7%**

| | | Target Text Issues | | | | | | Source Text Issues | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Issue counts | | | Penalties | | | Issue counts | | | Penalties | | |
| Issue Type | Weight (default=1) | Minor | Major | Critical | Raw | Adj. | Target Subscore | Minor | Major | Critical | Raw | Adj. | Source Subscore |
| **Accuracy** | | | | | | | | | | | | | |
| Terminology | 1.0 | 1 | | | 1 | 1 | 99.7% | | | | | | |
| Mistranslation | 1.0 | 1 | | | 1 | 1 | 99.7% | | | | | | |
| Omission | 1.0 | | | | 0 | 0 | 100% | | | | | | |
| Untranslated | 1.0 | | | | 0 | 0 | 100% | | | | | | |
| *Accuracy Subtotal* | | | | | — | 2 | 99.3% | | | | | | |
| **Fluency** | | | | | | | | | | | | | |
| Unintelligible | 1.0 | | | | 0 | 0 | 100% | 1 | | | 1 | 1 | 99.5% |
| **Content** | | | | | | | | | | | | | |
| Inconsistency | 1.0 | 4 | 2 | 0 | 14 | 14 | 95.3% | 1 | | | 1 | 1 | 99.5% |
| Register | 1.0 | | | | 0 | 0 | 100% | | | | 0 | 0 | 100% |
| **Mechanical** | | | | | | | | | | | | | |
| Spelling | 1.0 | | | | 0 | 0 | 100% | | | | 0 | 0 | 100% |
| Grammar | 1.0 | | | | 0 | 0 | 100% | | | | 0 | 0 | 100% |
| *Fluency Subtotal* | | | | | — | 14 | 95.3% | | | | — | 2 | 99% |
| **Verity** | | | | | | | | | | | | | |
| Legal requirements | 1.0 | | | | 0 | 0 | 100% | | | | 0 | 0 | 100% |
| *Verity Subtotal* | | | | | — | 0 | 100% | | | | — | 0 | 100% |
| **Total** | | | | | — | 16 | 94.7% | | | | — | 2 | 99% |

Save Metric

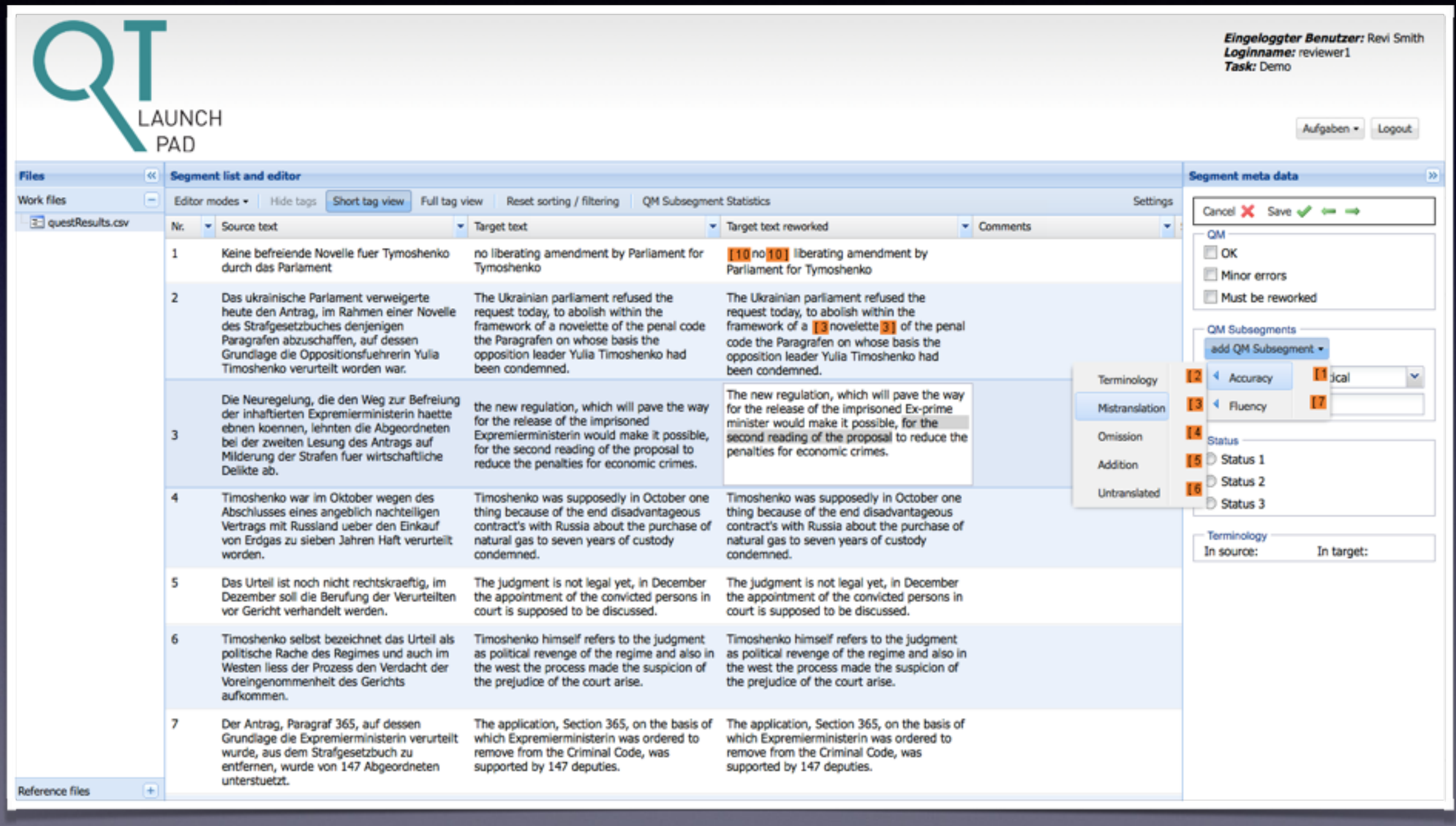
http://www.qt21.eu/MQM

# Ergonomic Scorecard




http://www.qt21.eu/MQM

# translate5




**DEMO: http://www.translate5.net**

# Currently clarifying the (complementary) relationship between MQM and the TAUS DQF Error Typology

QT LAUNCH PAD

Learn more at

http://www.qt21.eu

LAUNCH
PAD

# Questions?

QT LAUNCH PAD