

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.



PROBLEMS IN ASSESSING QUALITY

95% of professionally translated content at one major LSP is **never evaluated** for quality.

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

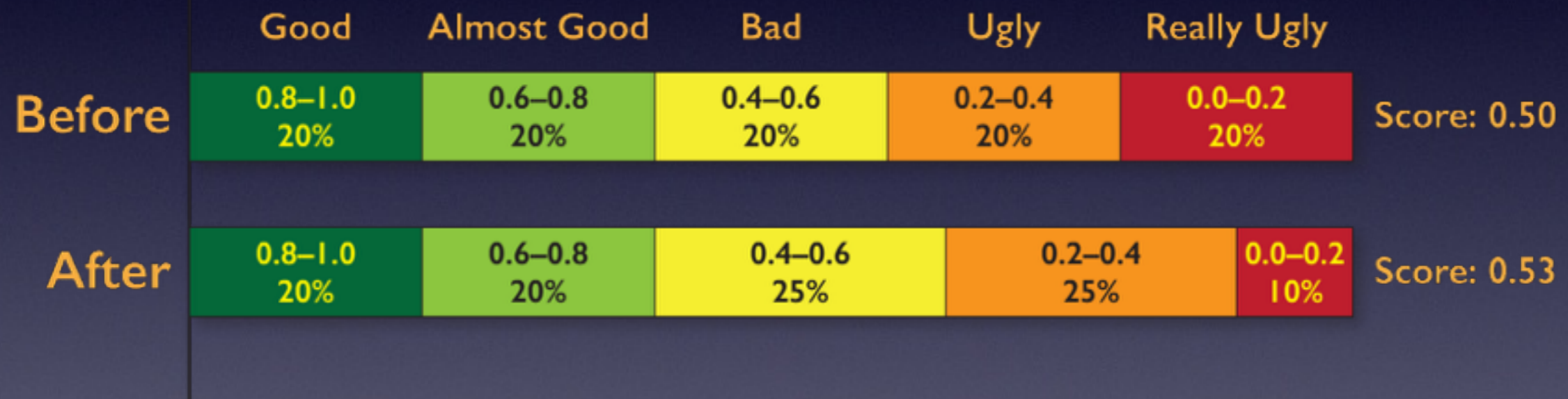
“I know it when I see it”

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

MT methods require
reference translations:
Cannot be used for
production purposes

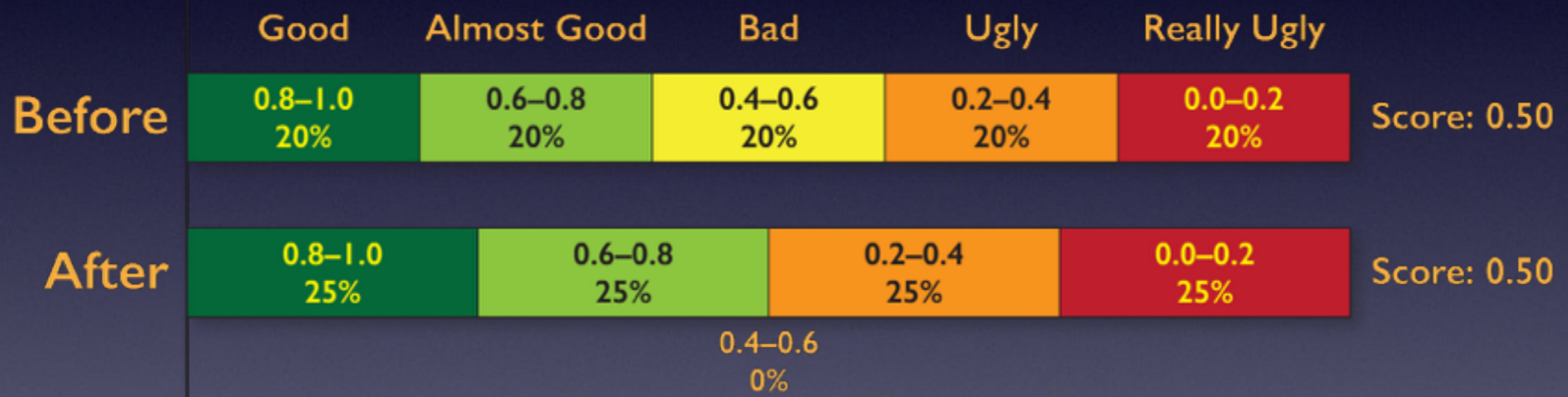
Change the reference
translation(s) and
the score changes

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

**Most sampling today
is random but
errors are not random**

Wait a minute...

**What do you mean by
quality?**

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?

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

Multidimensional Quality Metrics

MQM

Why not use a
single metric
everyone shares?

OK.

Which one?

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

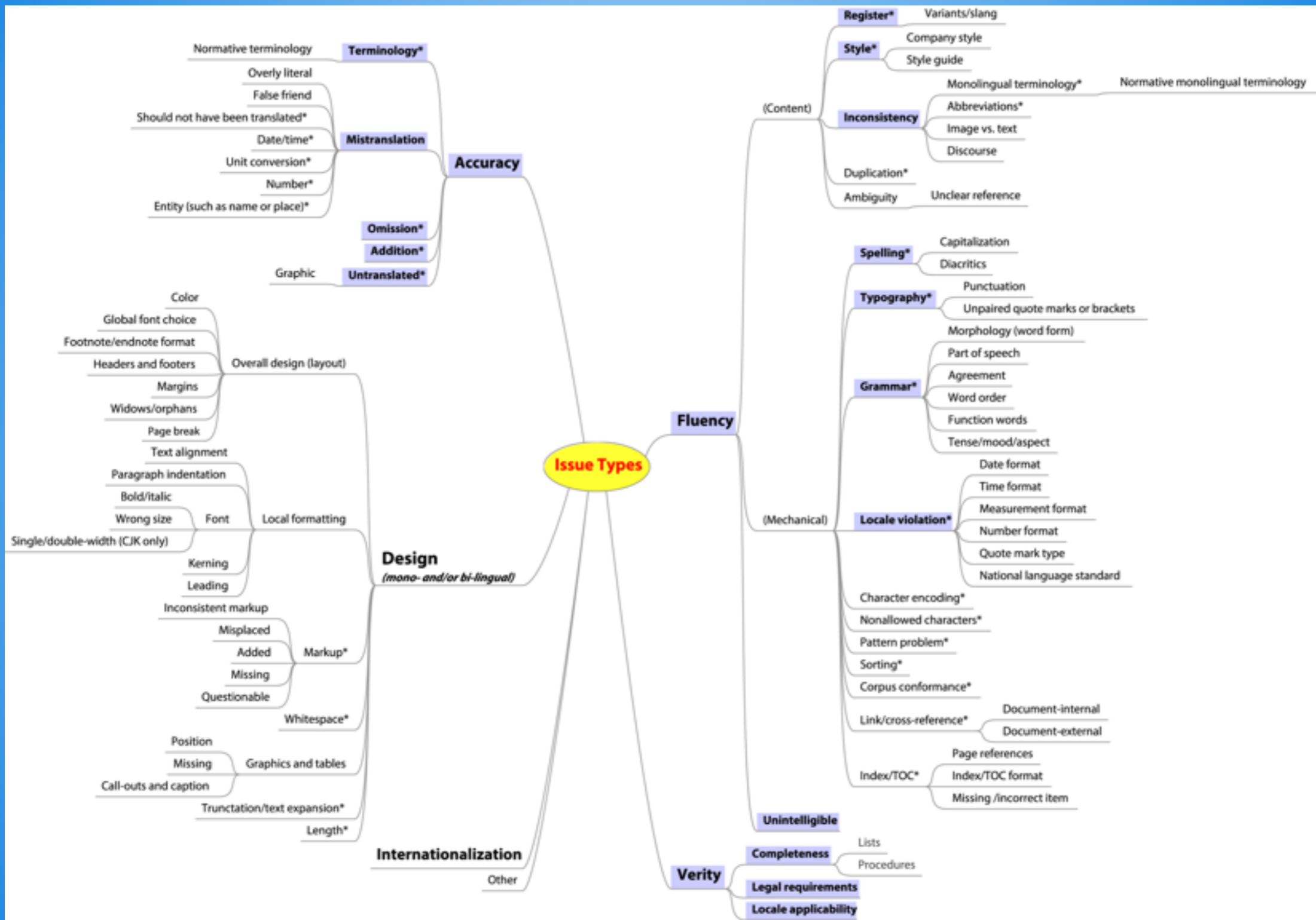
All of them disagree* about
what is **important to**
quality

*The only thing they agree on is terminology

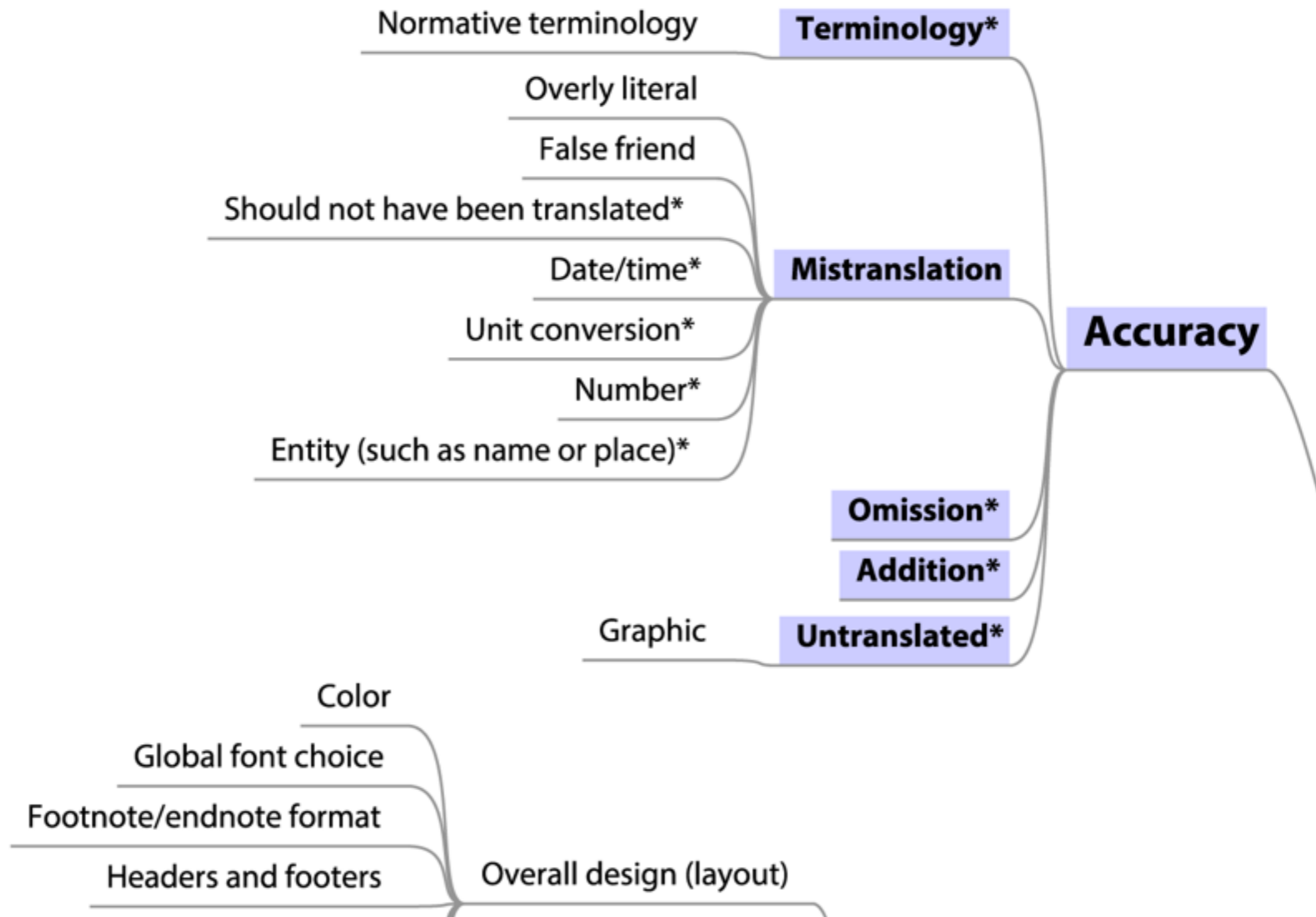
(Probably because there is
no single set of criteria
that applies to all kinds of
translation)

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

MQM provides a **catalog of issue types** suitable for various tasks

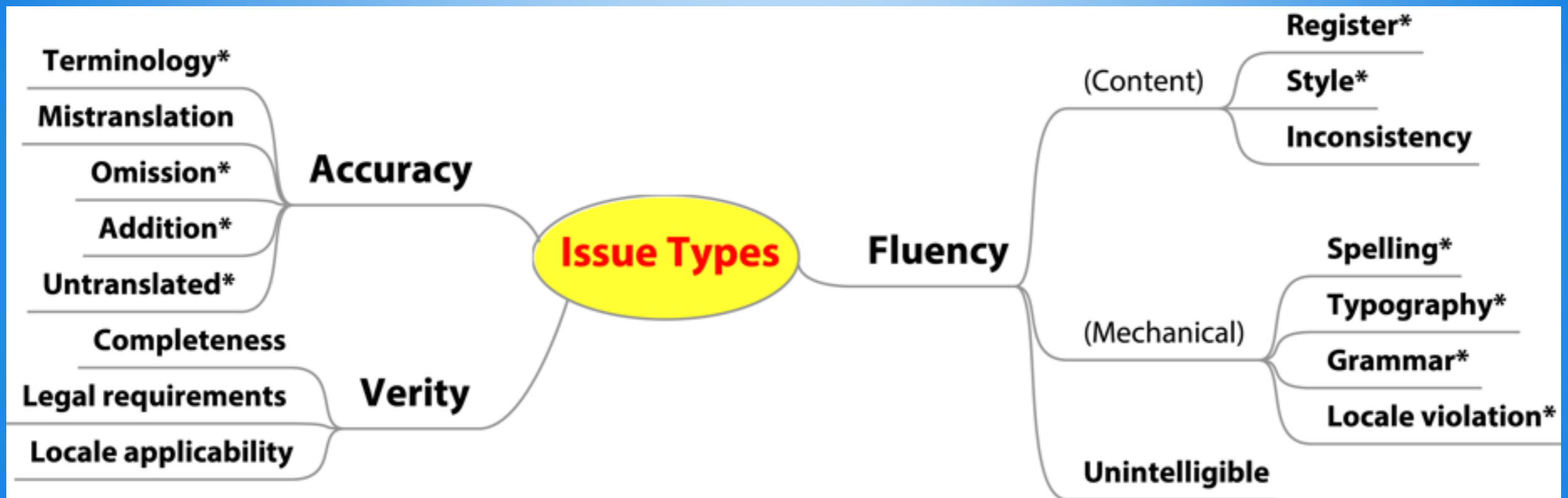


The "full" MQM ("Spaghetti")



Zooming into the "full" MQM

Wait! Weren't we trying to
improve things?



The MQM Core

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

Accuracy and Fluency

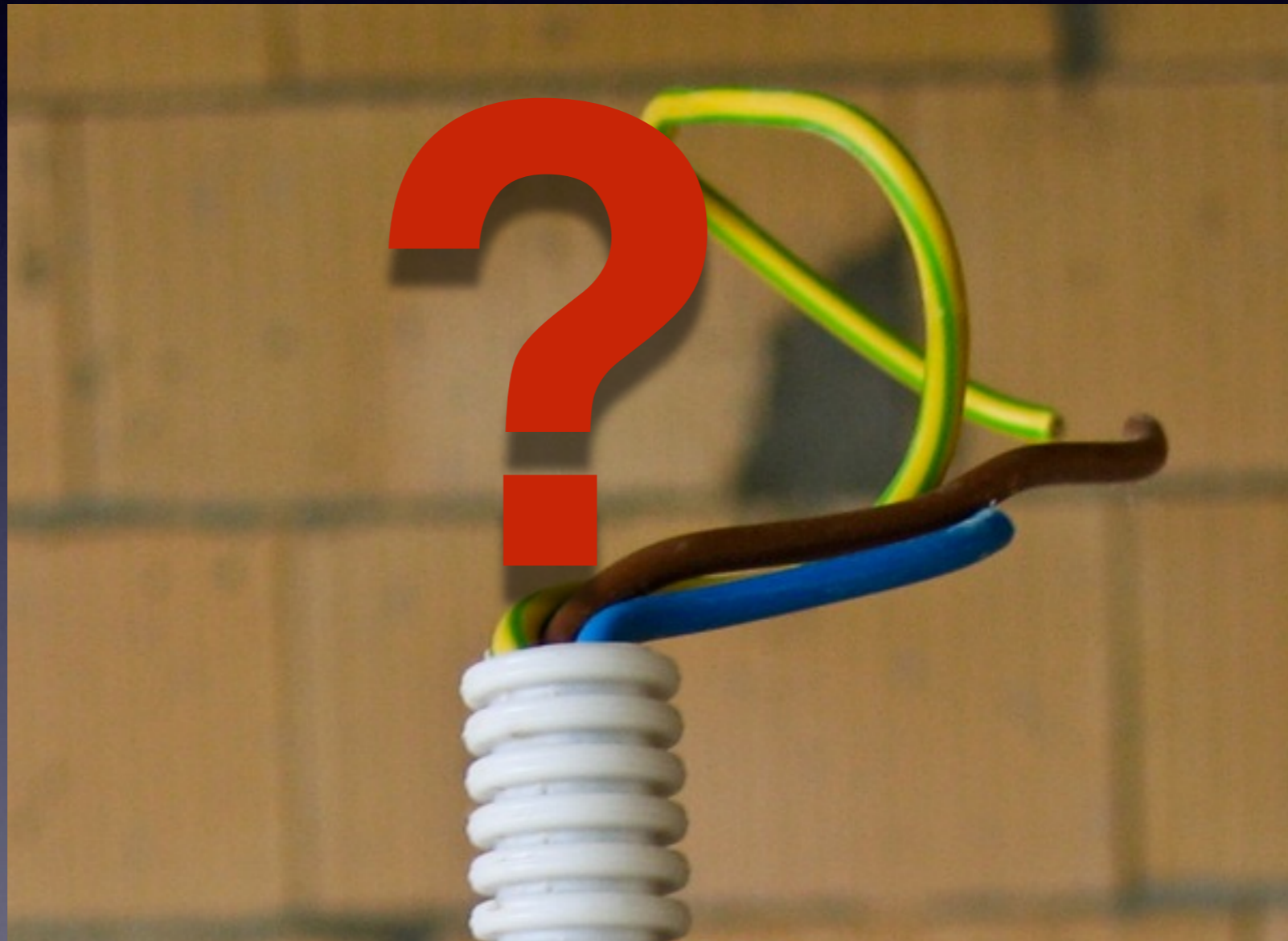
What's Verity?

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

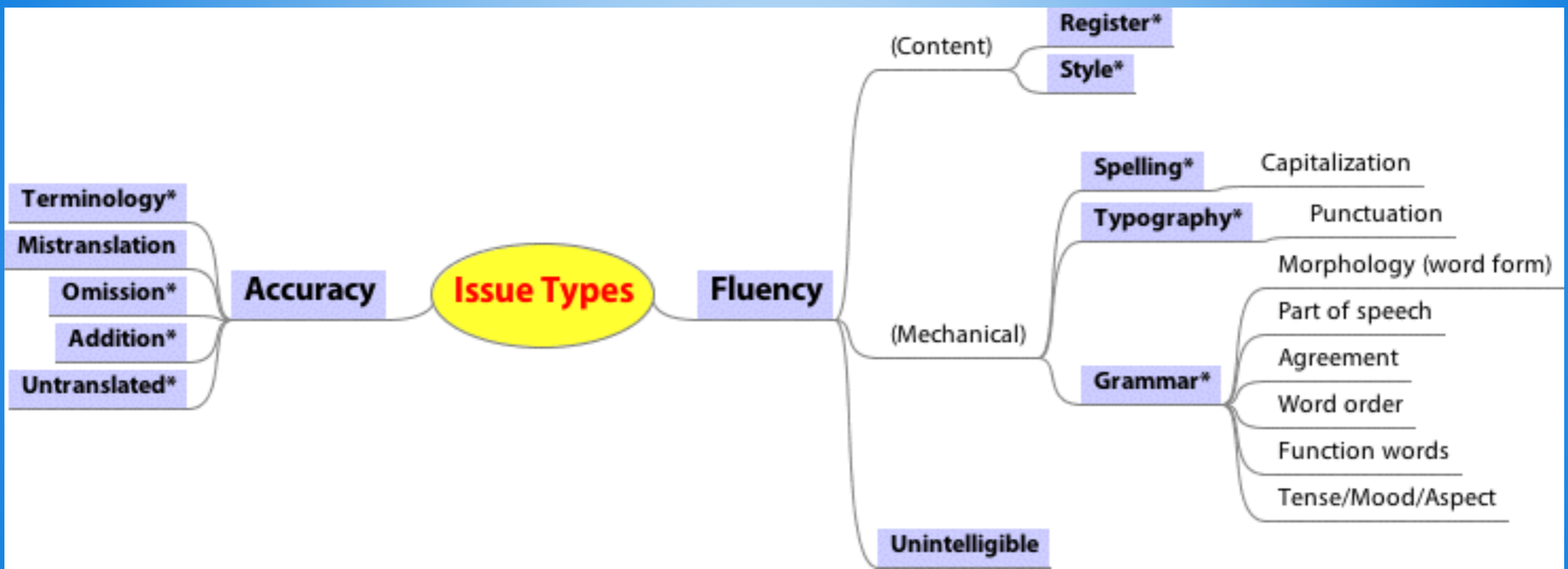
Locate the ground wire (bare copper)...



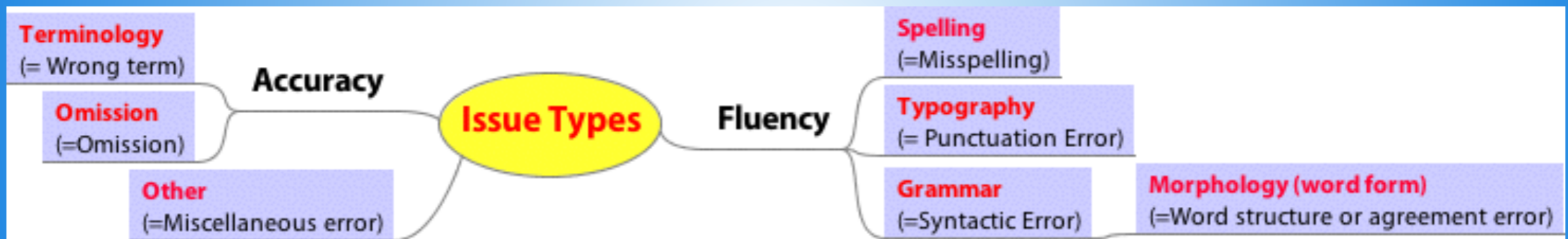
Finden Sie das Erdungskabel (**blankes Kupfer**)



You don't use all of MQM
(or its core):
you **use the**
parts you need.



MQM for MT Diagnostics



SAE J2450

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

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

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

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



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:	<input type="text" value="Russian"/>	<input type="text" value="RUSSIAN FEDERATION"/>
target	<input type="text" value="English"/>	<input type="text" value="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)

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 Save Metric

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%						
Verity													
Legal requirements	1.0				0	0	100%				0	0	100%
Verity Subtotal					—	0	100%						
Total					—	16	94.7%						

Show Metric

Ergonomic Scorecard

Word Counts Word Count (Source): <input type="text" value="200"/> Word Count (Target): <input type="text" value="200"/>	Clear Scorecard <input type="button" value="Clear Scorecard"/>	Upload a .csv Formatted Translation File <input type="button" value="Choose File"/> no file selected <input type="button" value="Submit"/>	Save status Synced: <input checked="" type="checkbox"/>
--	--	---	---

Score: 97.5%	Source <input type="text" value="quelle"/>	Target <input type="text" value="ziel"/>	Comments <input type="button" value="Add a Comment..."/> <input type="button" value="View Comments"/>
← Previous	Segment 1 of 11		Next →

Accuracy	Fluency				Verity	
Accuracy (General) (0) TARGET	Fluency (General) (10) SOURCE	Fluency (General) (15) TARGET	Mechanical (General)(5) SOURCE	Mechanical (General)(0) TARGET	Verity(General)(0) SOURCE	Verity(General)(0) TARGET
Terminology (0) TARGET	Content (General) (0) SOURCE	Content (General) (0) TARGET	Spelling(0) SOURCE	Spelling(0) TARGET	Legal Requirements(0) SOURCE	Legal Requirements(0) TARGET
Mistranslation (0) TARGET	Inconsistency(0) SOURCE	Inconsistency(0) TARGET	Grammar(0) SOURCE	Grammar(5) TARGET		
Omission (0) TARGET	Register(0) SOURCE	Register(0) TARGET	Unintelligible(0) SOURCE	Unintelligible(0) TARGET		
Untranslated (0) TARGET						

translate5

The screenshot displays the translate5 web interface. At the top left is the 'QT LAUNCH PAD' logo. The top right shows user information: 'Eingeloggter Benutzer: Revi Smith', 'Loginname: reviewer1', and 'Task: Demo'. Below this are 'Aufgaben' and 'Logout' buttons. The main area is titled 'Segment list and editor' and contains a table with columns: 'Nr.', 'Source text', 'Target text', 'Target text reworked', and 'Comments'. The table lists seven segments with German source text and English target text. A context menu is open over the third segment, listing 'Terminology', 'Mistranslation', 'Omission', 'Addition', and 'Untranslated'. The right sidebar, 'Segment meta data', includes 'Cancel', 'Save', and 'QM' checkboxes for 'OK', 'Minor errors', and 'Must be reworked'. It also features 'QM Subsegments' with an 'add QM Subsegment' button and a list of subsegments: 'Accuracy' (value: jcal), 'Fluency' (value: 7), and 'Status' (radio buttons for Status 1, 2, 3). A 'Terminology' section at the bottom of the sidebar has 'In source:' and 'In target:' fields.

Nr.	Source text	Target text	Target text reworked	Comments
1	Keine befreiende Novelle fuer Tymoshenko durch das Parlament	no liberating amendment by Parliament for Tymoshenko	[10 no 10] liberating amendment by Parliament for Tymoshenko	
2	Das ukrainische Parlament verweigerte heute den Antrag, im Rahmen einer Novelle des Strafgesetzbuches denjenigen Paragrafen abzuschaffen, auf dessen Grundlage die Oppositionsfuehrerin Yulia Tymoshenko verurteilt worden war.	The Ukrainian parliament refused the request today, to abolish within the framework of a novelette of the penal code the Paragrafen on whose basis the opposition leader Yulia Tymoshenko had been condemned.	The Ukrainian parliament refused the request today, to abolish within the framework of a [3 novelette 3] of the penal code the Paragrafen on whose basis the opposition leader Yulia Tymoshenko had been condemned.	
3	Die Neuregelung, die den Weg zur Befreiung der inhaftierten Expremierministerin haette ebnet koennen, lehnten die Abgeordneten bei der zweiten Lesung des Antrags auf Milderung der Strafen fuer wirtschaftliche Delikte ab.	the new regulation, which will pave the way for the release of the imprisoned Expremierministerin would make it possible, for the second reading of the proposal to reduce the penalties for economic crimes.	The new regulation, which will pave the way for the release of the imprisoned Ex-prime minister would make it possible, for the second reading of the proposal to reduce the penalties for economic crimes.	
4	Timoshenko war im Oktober wegen des Abschlusses eines angeblich nachteiligen Vertrags mit Russland ueber den Einkauf von Erdgas zu sieben Jahren Haft verurteilt worden.	Timoshenko was supposedly in October one thing because of the end disadvantageous contract's with Russia about the purchase of natural gas to seven years of custody condemned.	Timoshenko was supposedly in October one thing because of the end disadvantageous contract's with Russia about the purchase of natural gas to seven years of custody condemned.	
5	Das Urteil ist noch nicht rechtskraeftig, im Dezember soll die Berufung der Verurteilten vor Gericht verhandelt werden.	The judgment is not legal yet, in December the appointment of the convicted persons in court is supposed to be discussed.	The judgment is not legal yet, in December the appointment of the convicted persons in court is supposed to be discussed.	
6	Timoshenko selbst bezeichnet das Urteil als politische Rache des Regimes und auch im Westen liess der Prozess den Verdacht der Voreingenommenheit des Gerichts aufkommen.	Timoshenko himself refers to the judgment as political revenge of the regime and also in the west the process made the suspicion of the prejudice of the court arise.	Timoshenko himself refers to the judgment as political revenge of the regime and also in the west the process made the suspicion of the prejudice of the court arise.	
7	Der Antrag, Paragraf 365, auf dessen Grundlage die Expremierministerin verurteilt wurde, aus dem Strafgesetzbuch zu entfernen, wurde von 147 Abgeordneten unterstuetzt.	The application, Section 365, on the basis of which Expremierministerin was ordered to remove from the Criminal Code, was supported by 147 deputies.	The application, Section 365, on the basis of which Expremierministerin was ordered to remove from the Criminal Code, was supported by 147 deputies.	

DEMO: <http://www.translate5.net>

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

Learn more at
<http://www.qt21.eu>

Questions?