ANNOTVIA AI

Arabic translation annotation sample

This short example shows how we would review an English translation against an Egyptian Arabic source, document errors, and deliver a corrected version. Every decision can be traced back to a specific source segment.

Demonstration disclosure: Created for this request with AI assistance. The dialogue and candidate translations are fictional; no client material or personal recordings are included. This is not a previous production delivery or a claim of independent human review.

Source transcript and candidate translation

Sample ID: ANV-KUDO-DEMO-001 | Source: Egyptian Arabic (ar-EG) | Target: English (en) | Rubric: v1.0
Scenario: A fictional project handover. This is a written source, not an audio transcription. Speaker labels are scripted; timestamps, ASR accuracy and acoustic quality are not assessed.

S01 Speaker A

إحنا مش هنبدأ التسجيل قبل ما كل المشاركين يوافقوا.

Candidate: We will start recording before all participants agree.

S02 Speaker B

تمام، الاجتماع اتنقل من الساعة تلاتة للساعة خمسة بتوقيت القاهرة.

Candidate: Okay, the meeting has moved from 3 p.m. to 4 p.m. Cairo time.

S03 Speaker A

ابعت النسخة النهائية يوم الخميس، بس لو المراجعة خلصت.

Candidate: Send the final version on Thursday.

S04 Speaker B

حاضر، هبعتلك الملف بعد المراجعة.

Candidate: Sure, I’ll send you the file after the review.

Completed annotation output

Scores below assess the candidate translations, not the corrected versions. A fluent sentence can still be wrong. The error severity determines the release decision; scores are never averaged to hide a critical error.

ID	Accuracy	Fluency	Severity	Decision
S01	1 / 5	5 / 5	Critical	Rework
S02	2 / 5	5 / 5	Critical	Rework
S03	3 / 5	5 / 5	Major	Rework
S04	5 / 5	5 / 5	None	Accept

S01 Negation reversed

Error label: ACCURACY_NEGATION | Target span: “will start recording before”

Evidence: “مش هنبدأ ... قبل ما ... يوافقوا” means recording must wait until everyone agrees. The candidate reverses that safeguard.

Final corrected translation: We won’t start recording until all participants agree.

S02 Time changed and detail added

Error labels: ACCURACY_TIME; UNSUPPORTED_ADDITION | Target span: “3 p.m. to 4 p.m.”

Evidence: “من الساعة تلاتة للساعة خمسة” means from three to five. The candidate changes five to four and adds p.m., which the written source does not specify. The wrong meeting time drives the Critical label.

Final corrected translation: Okay, the meeting has been moved from three to five, Cairo time.

S03 Condition omitted

Error label: OMISSION_CONDITION | Target span: missing clause after “Thursday”

Evidence: “بس لو المراجعة خلصت” makes delivery conditional on the review being finished. The candidate turns a conditional instruction into an unconditional one. Under this demo rubric, that is Major; a safety-critical condition would be Critical.

Final corrected translation: Send the final version on Thursday, but only if the review is finished.

S04 Meaning preserved

Error label: NONE | Target span: none

Evidence: The acknowledgment, file delivery and “after the review” sequence are all preserved. “Sure” is a natural rendering of “حاضر” in this context.

Final corrected translation: Sure, I’ll send you the file after the review.

Delivery outcome: 1 candidate accepted unchanged; 3 require correction. The corrected translations above form the completed demonstration output. Production release would still require the agreed independent review and sign-off.

Scoring rubric

Project-specific demonstration rubric, v1.0. These rules are illustrative and would be calibrated with KUDO before production. Accuracy measures faithfulness to the source; fluency measures English readability independently.

Score	Accuracy	Fluency
5	All source meaning and material details preserved.	Natural, grammatical and easy to follow.
4	Minor imprecision; no material change in meaning.	Minor awkwardness; easy to understand.
3	Main message retained, but a material detail or condition is lost or changed.	Noticeable wording or grammar problems; meaning remains clear.
2	Major distortion of an essential detail; only part of the message is reliable.	Frequent problems make understanding difficult.
1	Core meaning reversed, unrelated or largely missing.	Largely unintelligible.

Critical: An error reverses a consent/safety instruction or changes a high-impact operational fact such as the meeting time. Major: A material meaning error without that critical consequence. Minor: A limited imprecision with no material meaning change. None: No identified error.

Decision rule: Accept only if Accuracy and Fluency are both at least 4 and no Major or Critical errors remain. Otherwise rework. If the source is genuinely ambiguous, flag it for adjudication rather than guessing. Use the highest severity when a segment contains multiple errors.

How we would check reviewer quality

Before starting, we would agree a small set of reference examples with KUDO. Reviewers would score them independently, then discuss differences until the labels and severity boundaries are clear. Correcting a fluent but unfaithful translation would be part of qualification.

For a pilot, we propose a second reviewer checking every item without seeing the first reviewer’s scores. A lead reviewer would resolve disagreements against the source and rubric, recording the reason for each change. For larger batches, the review percentage would be agreed with KUDO, with all critical flags and disputed items reviewed.

We would track exact severity agreement, accuracy-score differences, missed critical errors and adjudication overturns for each reviewer. Repeated errors would trigger targeted feedback, recalibration and a recheck of affected work. These are proposed controls, not measured results from this demonstration.

The delivery record would retain segment ID, source, candidate, error span and label, severity, both scores, rationale, corrected translation, reviewer IDs and rubric version. Audio-specific fields would be added only when real audio is supplied.

Scope and sharing: This fictional text sample is intended for capability review. It contains no licensed audio dataset and makes no claim about available recording volume, participant consent or production-data rights. Any training/evaluation licence for future data would be agreed separately.

Annotvia AI | contact@annotvia.com | https://annotvia.com/