

Subtitling Across Borders: Inference, Relevance, and Audience Access in Global Nature Documentary Translation

Baharuddin Baharuddin

bahar@unram.ac.id

Universitas Mataram

Lalu Ali Wardana

Universitas Mataram

Lalu Jaswadi Putera

Universitas Mataram

Desi Herayana

Universitas Mataram

I Gusti Agung Ayu Sasmitha Dewi Anggita Pramestya

Universitas Mataram

Alia Fitrida Camila

Universitas Mataram

Auliya Rahma

Universitas Mataram

M Sazwan

Universitas Mataram

Duwi Wulandary

Universitas Mataram

Research Article

Keywords: subtitling, relevance, ACSA, audiovisual translation, explicature, presupposition

Posted Date: August 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10608600/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Subtitling a nature documentary for a global audience is not simply a matter of finding target-language equivalents for a source narration; it is an exercise in pragmatic inference under constraints of space, time, and channel. This article examines how the English-to-Indonesian subtitles of Netflix's *Our Planet: Migrations* manage the recovery of intended meaning across four episodes, using the Audience-Centred Subtitling Adequacy (ACSA) framework to assess subtitle quality along three interacting dimensions: Micro (readability and processing ease), Meso (pragmatic alignment of intent and stance), and Macro (global coherence and accessibility). Four groups of trainee translators, each working on one episode of the series, independently scored a combined 44 subtitle segments. Pooled results show consistently high adequacy (Micro $M = 4.52$, $SD = 0.70$; Meso $M = 4.43$, $SD = 0.93$; Macro $M = 4.52$, $SD = 0.82$, all on a 1–5 scale), but with the Meso dimension showing the greatest variability and, together with Macro, the lowest individual scores in the dataset. Qualitative analysis of the lowest-scoring segments shows that adequacy loss concentrates at points where a subtitle is locally accurate but narrows, domesticates, or destabilises the pragmatic stance and ecological register of the source narration. We argue that this pattern is consistent with a relevance-theoretic account of subtitling as constrained pragmatic inference, in which explicature, implicature, and presupposition must be calibrated to the addressee's cognitive environment, and that the ACSA framework offers a tractable, pedagogically useful operationalisation of this calibration for subtitler training and quality assessment.

1. Introduction

When an Indonesian-speaking viewer watches Netflix's *Our Planet: Migrations* with subtitles, they are not simply reading a translation of David Attenborough's English narration. They are performing an act of pragmatic inference: reconstructing, from a compressed and culturally repositioned textual trace, an interpretation that approximates – as closely as the communicative situation allows – what the original utterance was intended to convey, with its accompanying cognitive and affective effects. The subtitle, on this view, is not a code but an ostensive stimulus: evidence of a communicative intention that the viewer must inferentially enrich (Sperber and Wilson, 1995; Gutt, 2000).

This article takes that inferential dimension of subtitling as its object of study, and examines it through a structured quality-assessment exercise rather than through viewer reception data. Four groups of trainee translators, working independently as part of an audiovisual translation course, each took responsibility for one episode of *Our Planet: Migrations* (Episodes 1 through 4) and applied the Audience-Centred Subtitling Adequacy (ACSA) framework to score subtitle segments from their assigned episode, evaluating each segment along three dimensions: Micro (readability and processing ease for the viewer), Meso (the degree to which pragmatic intent, evaluative stance, and tone are preserved), and Macro (the segment's contribution to global narrative coherence and accessibility). This three-tier structure operationalises, in a form suitable for systematic scoring, the same theoretical distinctions – explicature, implicature, and presupposition – that relevance-theoretic accounts of translation treat as central to successful communication under constraint (Gutt, 2000; Carston, 2002).

We report the pooled results of this scoring exercise across 44 independently assessed segments, examine the pattern of variation across the three ACSA dimensions, and analyse in detail the small number of segments that received the lowest scores. We argue that these low-scoring cases are not randomly distributed errors but cluster around a specific and theoretically diagnostic failure mode: subtitles that are locally accurate at the level of individual lexical choice but that narrow, domesticate, or otherwise destabilise the pragmatic stance, register, or presupposed cognitive environment of the source narration. This pattern is precisely what a relevance-theoretic account of subtitling predicts will be the hardest dimension of subtitle quality to achieve, and the easiest to overlook in conventional fidelity-focused translation training.

The remainder of the article is organised as follows. Section 2 develops the theoretical link between relevance theory and the three-tier ACSA framework. Section 3 describes the scoring exercise undertaken by the four trainee groups. Section 4 reports pooled quantitative results and presents detailed qualitative analysis of the lowest-scoring segments from each affected group, alongside contrasting evidence from the highest-scoring group. Section 5 discusses the implications of this pattern for pragmatic theory and for subtitler training, and Section 6 concludes with directions for future work that would combine expert pragmatic assessment of the kind reported here with viewer-facing reception data.

1.1 Our Planet: Migrations as a case study

Our Planet: Migrations is part of the Our Planet documentary franchise produced by Netflix in partnership with the World Wide Fund for Nature, narrated by Sir David Attenborough. Like its predecessor series, it combines high-density scientific narration – species names, ecological mechanisms, climatic causation – with a lyrical, affectively marked narrative register intended to cultivate wonder, urgency, and environmental concern in the viewer. The four episodes examined in this study cover a wide range of migratory phenomena: the documentary's opening framing of human and animal migration together with African buffalo and lion predation in the Kalahari (Episode 1); Laysan albatross fledging, tiger shark predation, and honeybee colony behaviour (Episode 2); human population growth, rainforest biodiversity, and lowland gorilla social structure (Episode 3); and Asian elephant migration through Kunming alongside gentoo penguin breeding behaviour and further Arctic and North American material (Episode 4). This combination of dense scientific content, affective narrative framing, and rapid shifts of geographic and species focus makes the series a demanding test case for subtitle quality across all three ACSA dimensions simultaneously. The present study extends a prior strategy-level analysis of the same documentary franchise's YouTube caption track (Baharuddin et al., 2025), moving from a focus on translation-strategy distribution to a pragmatically oriented, audience-centred adequacy assessment of the Netflix-distributed series.

It is worth noting that the four episodes, while united by the series' migration theme, differ considerably in their rhetorical structure. Episode 1's opening segments are explicitly reflective and scene-setting, establishing the documentary's human-and-animal migration framing in short, aphoristic lines before

transitioning into observational wildlife narration; Episode 4's segments, by contrast, are organised around explicit causal chains – drought causing migration, sunlight causing ice-melt causing breeding behaviour – that place a premium on the subtitle's ability to preserve logical connectives and causal sequencing across sentence boundaries. This structural variation across episodes is relevant to interpreting the between-group differences reported in Section 4: a sequence built from short, standalone aphorisms imposes different pragmatic demands on the subtitler than a sequence built from extended causal argument, even before any difference in translator skill or rubric stringency is considered.

2. Theoretical background

2.1 Relevance and the subtitle as ostensive stimulus

Relevance theory (Sperber and Wilson, 1995; Wilson and Sperber, 2004) holds that communication is successful when an utterance achieves optimal relevance for its addressee: when it yields adequate cognitive effects for no gratuitous processing effort. Applied to subtitling, this principle has a distinctive consequence. The subtitle is produced under severe effort-side constraints – limited reading time, limited line length, competition with the visual channel for attention – that are largely absent from ordinary spoken or written communication. The empirical psycholinguistics of subtitle reading bears this out directly: eye-tracking research since d'Ydewalle et al. (1987) has established that viewers allocate only a fraction of available subtitle display time to reading the text itself, dividing attention between verbal and visual channels in ways a purely text-based model of reading effort does not capture (see also Perego et al., 2010; Kruger and Steyn, 2014). A subtitle that reproduces the full propositional content of the source narration, word for word, may therefore fail to achieve relevance not because it is inaccurate but because it imposes processing costs the viewing situation cannot sustain.

This reframes the central design problem of subtitling, and maps directly onto the three-tier structure of the ACSA framework used in this study. The Micro dimension of ACSA corresponds to the effort side of the relevance calculation: a segment scores well on Micro when its explicature – its explicitly stated propositional content – can be processed by the viewer without unnecessary cost, regardless of whether it preserves every nuance of the source. The Meso dimension corresponds to the effect side, specifically as it concerns pragmatic stance: a segment scores well on Meso when the evaluative, attitudinal, and illocutionary force of the original utterance – its implicature, in relevance-theoretic terms – survives translation, even where this requires departing from literal lexical correspondence. The Macro dimension extends this effect-side evaluation to the level of discourse coherence and presupposition management: a segment scores well on Macro when it does not require the viewer to possess background assumptions – about species, geography, or the documentary's implicit conservation framing – that they may not have, and when it sustains rather than disrupts the narrative's overall coherence.

This three-way mapping also predicts a specific kind of interaction that a purely additive view of subtitle quality – in which a good subtitle simply scores well on every dimension at once – would not anticipate.

Because explicature and presupposition management both compete for the same limited textual space, a translator who increases explicitness to manage a presupposition risk (spelling out a term the source narration leaves implicit) necessarily increases processing cost, and a translator who compresses to minimise processing cost risks leaving a presupposition unaddressed. The ACSA framework's separation of Micro from Macro allows this trade-off to be observed empirically rather than assumed away, and Section 4.3 reports a clear instance of exactly this trade-off operating in the dataset.

2.2 Explicature, implicature, and presupposition under subtitling constraints

Three pragmatic notions organise our analysis. Explicature concerns what a subtitle makes linguistically explicit, as opposed to leaving to inference; subtitlers face a continual choice between spelling out a piece of information and leaving it to be inferred from the image or omitted as not worth the processing cost. Implicature – content conveyed without being said outright, classically analysed in terms of the Gricean (1975) cooperative principle and its maxims, and developed further by Levinson (2000) in terms of default, generalised inferential patterns – concerns content a subtitle is designed to convey without stating it outright: a connotation, an evaluative stance, an emotional colouring achieved through word choice rather than propositional content. Presupposition concerns the background assumptions a subtitle treats as already available to its addressee; a subtitle that uses an unglossed technical term, or that assumes familiarity with a particular framing of human–nature relations, is presupposing a cognitive environment that may or may not match the actual viewer's. This kind of presupposition about an addressee's social or cognitive position is not unique to subtitling: comparable assumptions have been documented in address-form choice in postcolonial Sasak communities, where the selection of a particular form of address presupposes, and thereby helps constitute, the addressee's social position relative to the speaker (Yusra et al., 2025).

A brief example from the Episode 1 dataset illustrates the interaction of these three dimensions, and anticipates a pattern discussed at length in Section 4. The English line “At any given moment, billions of animals are on the move” was rendered as “Di masa-masa tertentu, miliaran satwa bermigrasi” (literally, “at certain times”). The subtitle is fully processable – it scored well on Micro – but “tertentu” (certain, particular) implicates selectivity and intermittency, where the source's “any given moment” explicature asserts constancy. This is a Meso-level failure: the subtitle has not misstated a fact, but it has altered the implicated stance the original utterance was designed to convey, and Group 1 scored this segment's Meso dimension at 2 out of 5, the lowest score recorded anywhere in the dataset.

A related but more subtle case from the same sequence concerns the line “billions of animals are on the move,” rendered as “miliaran satwa bermigrasi” (“billions of creatures migrate”). The Indonesian verb “bermigrasi” is precise and the segment scored well on Micro and Meso (4 out of 5 each), but Group 1's own Macro rationale observes that “‘on the move’ is broader than ‘bermigrasi’; some species move non-migratorily,” awarding a Macro score of 3. The source explicature is deliberately broader than strict migration – covering any movement, migratory or otherwise – because the documentary's opening

argument is precisely that movement in general, not migration specifically, is what sustains planetary health; narrowing this to “bermigrasi” quietly forecloses part of the claim the documentary is building toward. Comparable lexical-choice failures, in which a single word choice imports an unintended presupposed meaning, have been documented in other short, space-constrained translation contexts, such as bilingual tourist signage, where the rendering of “periodically” as the Indonesian equivalent of “once in a while” rather than “regularly” reversed the intended public-health instruction (Baharuddin et al., 2024). This case shows that presupposition and coherence risk can arise even from translation choices that are, at the sentence level, both accurate and idiomatic, when those choices interact with an argument being constructed gradually across multiple subtitle units rather than within any single one.

2.3 ACSA as an operationalisation of pragmatic adequacy

The Audience-Centred Subtitling Adequacy (ACSA) framework, developed by the first author as part of an ongoing research and teaching programme in audiovisual translation (Baharuddin, unpublished), provides a structured, three-tier instrument for assessing subtitle quality that goes beyond binary correct/incorrect judgements of semantic equivalence. The version of the rubric applied in the present study scores the three tiers on a 1–5 scale and treats them as independently assessed dimensions; a related, more recent formulation of ACSA developed for other corpora in the same research programme uses a condensed 1–3 scale with an explicit hierarchical relationship between tiers. The two formulations share the same conceptual architecture and the same three-way distinction among readability, pragmatic alignment, and global coherence, but the present study reports results using the 1–5 scale and the independently scored design, since that is the version the assessing groups were trained on and applied. We treat the two formulations as purpose-differentiated instruments within a single research programme, in a manner broadly analogous to Skopos theory’s rejection of a single universal translation standard in favour of adequacy relative to a translation’s specific communicative purpose (Vermeer, 1989): the 1–5 independently scored design is suited to fine-grained, comparative, segment-level diagnostic work of the kind reported here, while the condensed, hierarchical 1–3 design is suited to characterising a whole subtitle track’s overall communicative adequacy as a single corpus-level verdict. Its Micro tier targets the cognitive-processing dimension of subtitle reception: readability, line length, and the absence of unnecessary parsing difficulty. Its Meso tier targets pragmatic alignment: whether the illocutionary force, evaluative stance, and rhetorical effect of the source utterance are preserved in the target text, independent of literal lexical correspondence. Its Macro tier targets discourse-level and accessibility outcomes: whether a segment, considered as part of the unfolding documentary narrative, supports or undermines the viewer’s capacity to follow the argument and ecological claims being advanced.

Because each tier targets a distinct locus of potential communicative failure, ACSA allows assessors to identify cases – of exactly the kind illustrated above – where a subtitle succeeds at one tier while failing at another. This granularity is the framework’s principal methodological advantage for the present study: it allows us to ask not simply whether the subtitles of *Our Planet: Migrations* are good or bad, but specifically where, in pragmatic terms, adequacy is most at risk, and to test the relevance-theoretic

prediction that effect-side dimensions (Meso, Macro) will show more variable and more frequently compromised outcomes than the effort-side dimension (Micro), given the genre's combination of dense scientific content and affectively loaded narration.

3. Method

Subtitle quality was assessed using the ACSA framework by four independent groups of trainee translators enrolled in an audiovisual translation course, with each group assigned responsibility for one episode of *Our Planet: Migrations*. Each group selected a set of subtitle segments from its assigned episode, transcribed the English source narration and the corresponding Indonesian subtitle for each segment, and scored each segment along the three ACSA dimensions on a 1–5 scale, accompanied by a written justification for each score, consistent with the view that translation quality judgement is properly understood as a matter of negotiated agreement among assessors rather than as the application of a fixed external metric (Lauscher, 2000). The full ACSA rubric used by the groups, including the anchor descriptions for each score point on all three dimensions, is provided as Supplementary Material. Scoring was conducted by group consensus following group discussion of each segment, rather than by individual raters working in isolation; this reflects standard practice in translation quality assessment training, where the verbalised justification accompanying a score is treated as pedagogically as important as the score itself.

Group 1 scored 14 segments from Episode 1, covering the documentary's opening framing of human and animal migration and a sequence on Kalahari buffalo and lion predation. Group 2 scored 10 segments from Episode 2, covering Laysan albatross fledging, tiger shark predation, and honeybee colony behaviour. Group 3 scored 10 segments from Episode 3, covering human population growth, rainforest biodiversity, and lowland gorilla social structure. Group 4 scored 10 segments from Episode 4, covering Asian elephant migration through Kunming, gentoo penguin breeding behaviour, and Arctic and North American material. In total, 44 segments were independently scored across the four groups, each segment receiving a Micro, a Meso, and a Macro score together with a written rationale.

We note two methodological features of this design that distinguish it from viewer-reception studies. First, the assessors in this study are not naive viewers but trainee translators applying an explicit analytic framework; the scores reported here therefore measure expert-judged pragmatic adequacy rather than directly measured viewer comprehension or engagement, and should be interpreted accordingly. Second, because each episode was scored by a single group rather than by a shared pool of raters working across all material, the pooled results reported in Section 4 reflect between-episode as well as within-segment variation, and we are correspondingly cautious about treating differences between groups as indicating differences in subtitle quality between episodes as opposed to differences in group judgement or episode difficulty.

4. Findings

4.1 Pooled ACSA scores

Across the 44 segments scored by the four groups, mean scores were high on all three ACSA dimensions: Micro $M = 4.52$ ($SD = 0.70$), Meso $M = 4.43$ ($SD = 0.93$), and Macro $M = 4.52$ ($SD = 0.82$), on the 1–5 scale. The pooled mean across all three dimensions was 4.49. Of the three dimensions, Meso showed both the lowest mean and, more strikingly, the highest standard deviation, indicating that pragmatic-stance preservation was the dimension on which subtitle quality varied most across segments, even though the average level of adequacy remained high. The lowest individual scores recorded anywhere in the dataset were a Meso score of 2 (the “any given moment” segment discussed in Section 2.2, from Episode 1) and a Macro score of 2 (discussed in Section 4.3 below, from Episode 3); no Micro score below 3 was recorded in any group’s assessment.

The overall distribution of scores across the dataset was strongly right-skewed on all three dimensions: a 5-out-of-5 score was the single most common outcome for Micro (28 of 44 segments, 64 per cent), Meso (29 of 44 segments, 66 per cent), and Macro (31 of 44 segments, 70 per cent). Scores of 2 occurred only three times across the entire dataset – once on Meso and once on Macro for the segments discussed in Sections 4.2 and 4.3, plus one further Meso score of 2 – and no score of 1 was recorded anywhere. This distribution is consistent with a dataset in which the great majority of translation decisions are pragmatically unproblematic, against a smaller number of segments where a specific, identifiable translation choice produces a clear and locatable adequacy cost; it is not consistent with a pattern of generally mediocre or inconsistent translation quality.

Group-level means showed some variation: Group 4 (Episode 4) recorded a perfect mean of 5.00 across all three dimensions, Group 2 (Episode 2) recorded an overall mean of 4.67 (Micro 4.60, Meso 4.70, Macro 4.70), Group 3 (Episode 3) recorded an overall mean of 4.27 (Micro 4.10, Meso 4.10, Macro 4.60), and Group 1 (Episode 1) recorded the lowest overall mean at 4.17 (Micro 4.43, Meso 4.07, Macro 4.00). We treat this group-level variation cautiously, for the reasons noted in Section 3: it may reflect genuine differences in subtitle quality across episodes, differences in the difficulty or density of the source material each group worked with, or differences in the stringency of each group’s application of the ACSA rubric. What is consistent across three of the four groups, despite this variation in overall level, is that Meso was the lowest-scoring or joint-lowest-scoring dimension; only Group 2 showed Meso scoring above its corresponding Micro score.

Group 4’s uniformly perfect scores merit closer attention, since a ceiling effect of this kind risks being uninformative on its own. Examination of Group 4’s rationales shows a consistent emphasis on the preservation of causal connectives across the Indonesian subtitles: rationales repeatedly credit segments for successfully conveying chains such as “drought → forced migration → search for new home” or “sunlight → warmth → ice melts → penguins breed,” explicitly noting that “logika kausal narasi terbangun dengan baik” (the narrative’s causal logic is well established) as the basis for high Meso and Macro scores. This suggests that Episode 4’s causally structured narration – noted in Section 1.1 – may have been comparatively easier to render with high pragmatic adequacy than Episode 1’s more

aphoristic opening sequence, simply because causal connectives of the kind Episode 4 relies on (“forced by,” “so that,” “because”) form part of the small, well-defined inventory of cohesive devices that Halliday and Hasan (1976) identify as carrying logical relations between clauses with comparatively little cross-linguistic slippage, in contrast to the implicit scope and stance distinctions – constancy versus selectivity, ecological versus domestic register – that proved costly in Episode 1 and that depend on lexical and pragmatic choices with no equally compact formal inventory.

A further consideration is that the four groups worked concurrently, during the same teaching period, without access to one another’s scoring decisions or rationales; the consistency we report in Section 5.1 across three of the four groups’ relative Micro–Meso–Macro ranking therefore emerged independently rather than through any shared discussion or calibration exercise between groups. We regard this as a modest point in favour of the pattern’s robustness, though it does not substitute for the kind of formal inter-rater reliability testing discussed as a limitation in Section 5.3.

4.2 The concentration of adequacy loss at the Meso tier

Qualitative examination of the written rationales accompanying lower-scoring segments shows a consistent pattern: Meso-tier score reductions are concentrated in cases where the Indonesian subtitle is lexically and syntactically unremarkable – sometimes scoring a full 5 on Micro – but narrows, domesticates, or shifts the evaluative register of the source. The “any given moment”/“tertentu” case discussed in Section 2.2 is the clearest example: Group 1’s own rationale notes that “‘any given moment’ implies constancy; ‘tertentu’ implies selectivity – stance shifted,” explicitly identifying the failure as one of pragmatic stance rather than propositional content.

A second example from Group 1’s Episode 1 dataset concerns the line “...or to find a new home,” rendered as “atau mencari rumah baru” (“or to seek a new home,” in the sense of a household). The group’s own Meso and Macro rationales note that “‘rumah baru’ shifts the tone toward a human/domestic interpretation rather than the scientific/ecological tone expected in a documentary context,” and that the term “habitat” would have better preserved the documentary’s ecological register; both dimensions were scored 3 out of 5 for this segment, against a Micro score of 5. Here again, the subtitle is locally fluent and entirely comprehensible, but a single lexical choice imports a connotation – domestic, human-scaled dwelling – that is not present in, and arguably works against, the source narration’s scientific and ecological framing.

A related but distinct pattern appears in Group 3’s Episode 3 assessment of a segment describing a silverback gorilla’s size: “He’s as tall as a man, but twice the weight,” rendered with a faithful comparative structure in Indonesian. Here the Meso and Macro scores (2 out of 5 each, the lowest Macro score in the dataset) were reduced not because of a register shift but because, in the group’s judgement, the rhythm and phrasing of the Indonesian comparative construction blunted the dramatic, almost incredulous force that the English construction achieves through its brevity – illustrating that Meso- and Macro-level adequacy can be put at risk by prosodic and rhetorical factors as well as by lexical-semantic ones.

A further case of this kind, from Group 3's assessment of the documentary's opening line, 'For the very first time, we look back at our own planet,' rendered as 'Untuk pertama kalinya, kita melihat kembali planet kita sendiri,' shows the same dynamic operating at the very start of an episode rather than mid-sequence. The group's Meso rationale, while still awarding 4 out of 5, notes that 'nuansa kontemplatif dari frasa 'look back at' sedikit berkurang dalam terjemahan Indonesia' (the contemplative nuance of 'look back at' is somewhat diminished in the Indonesian translation): 'melihat kembali' (to look again, to look back) is a perfectly natural rendering, but it lacks the retrospective, almost elegiac quality that 'look back at' carries in the source, where the phrase frames the entire documentary as a meditation on humanity's changed relationship to the planet rather than a simple act of looking. That this loss occurs in the documentary's opening line, before any subsequent scientific or narrative content has been established, suggests that the pragmatic stance the series sets out to occupy can be subtly recalibrated from its very first subtitle.

A further illustration of this prosodic sensitivity, scored slightly less severely but still below ceiling, comes from Group 1's assessment of "Since then, we haven't stopped," rendered as "Sejak saat itu, kita belum pernah berhenti," which continues in the source directly into "traveling to every corner of the Earth." The group's Macro rationale (scored 4 out of 5, the only non-maximal score in this short opening sequence) identifies the comma placement in the Indonesian construction – effectively "stopped, exploring" rather than the more natural "stopped exploring" – as introducing a minor disfluency at exactly the clause boundary where the source narration's momentum is most rhetorically important. The cost here is smaller than in the silverback case, but the locus of the cost is the same: a feature of the target text invisible to a purely lexical check, detectable only by attending to how the subtitle is paced against the surrounding narrative.

4.3 Macro-level coherence and presupposition

Macro scores were generally high ($M = 4.52$) but showed a distinct pattern from Meso: where Meso reductions concentrated around stance and register, Macro reductions concentrated around presupposition and terminology. Group 3's assessment of the line introducing a silverback gorilla ("They're led by this silverback, the dominant male") scored only 3 out of 5 on Micro specifically because, in the group's judgement, "silverback" – retained as a loanword in the Indonesian subtitle – "may still be unfamiliar to a general audience," even though the term was judged accurate and the segment scored well (4–5) on Meso and Macro. This is, in relevance-theoretic terms, a presupposition risk: the subtitle assumes the viewer either already possesses, or can readily infer from context, the encyclopaedic knowledge that "silverback" denotes a dominant adult male gorilla – an assumption that may not hold for all viewers, particularly given that the term is a culturally specific English compound with no established Indonesian equivalent.

A further Macro-relevant case, from Group 2's Episode 2 assessment, concerns the segment "arriving just in time for the fledging season," rendered as "tiba tepat saat musim anak burung mulai terbang" (literally, "arriving just as the season when young birds begin to fly"). This segment received the lowest Micro score in Group 2's dataset (3 out of 5), the group's rationale noting the relative length and

processing cost of the paraphrase compared with the single technical term “fledging” in the source; Meso and Macro scores for the same segment, by contrast, remained high (4 out of 5 each), since the paraphrase strategy – explaining rather than borrowing the technical term – successfully avoided the presupposition risk illustrated by the “silverback” case above, at the cost of some processing economy. The contrast between these two cases is instructive: where Group 3’s “silverback” borrowing preserved economy at the cost of presupposition risk, Group 2’s “fledging season” paraphrase preserved accessibility at the cost of processing economy, illustrating that Micro and Macro adequacy can trade off against one another even when Meso adequacy is held constant.

5. Discussion

5.1 Subtitling as constrained pragmatic inference

The pattern of findings reported in Section 4 supports treating subtitle quality, even as assessed through expert scoring rather than viewer reception, as a problem of constrained pragmatic inference rather than primarily one of semantic correspondence. Across three of the four groups, the dimension most analogous to processing effort in relevance-theoretic terms (Micro) was the most consistently high-scoring and least variable, while the dimensions most analogous to communicative effect – stance preservation (Meso) and discourse coherence and presupposition management (Macro) – showed both lower means and substantially higher variability. This is broadly the pattern a relevance-theoretic account predicts for a genre combining dense scientific content with affectively loaded narration: achieving adequate explicature is comparatively tractable, but calibrating implicature and presupposition to an uncertain and culturally heterogeneous addressee is harder, and more prone to the kind of subtle, locally undetectable slippage illustrated by the “tertentu” and “rumah baru” cases in Section 4.2. The “fledging season” and “silverback” cases in Section 4.3 add a further nuance: even within the effect-side dimensions, Macro-level accessibility and Micro-level economy can be placed in direct tension by a single translation strategy, suggesting that subtitle adequacy is not simply effort versus effect but a small system of interacting trade-offs that a translator must navigate segment by segment.

The strongly right-skewed distribution of scores reported in Section 4.1 is itself theoretically informative. A relevance-theoretic account does not predict that pragmatic calibration will usually fail; on the contrary, optimal relevance is, by hypothesis, what competent communicators ordinarily achieve, and translators – like speakers, on Gutt’s (2000) extension of the comprehension procedure to translation – can be expected to satisfy it in the great majority of cases without special difficulty. What the account predicts, and what this dataset shows, is that failures will be concentrated at specific, identifiable points of difficulty – scope distinctions, register-bearing lexical choices, prosodic phrasing at clause boundaries – rather than distributed evenly across the text. A quality-assessment exercise that returned uniformly mediocre scores, or uniformly high variance unrelated to identifiable linguistic features, would sit less comfortably with this theoretical picture than the pattern actually observed here.

It is worth being precise about what this pattern does and does not show. We do not claim that the subtitles assessed in this study are, on the whole, of poor quality – the pooled means on all three ACSA dimensions exceed 4.4 out of 5, indicating generally strong adequacy. Our claim is the narrower one that within this generally high-adequacy dataset, the cases that depart most from the ceiling are not randomly distributed but cluster at points of pragmatic rather than semantic risk, and that a quality-assessment framework built solely around semantic correspondence would have been poorly placed to detect this pattern at all.

5.2 Implications for subtitler training and the ACSA framework

These findings have direct implications for how subtitler training addresses pragmatic competence. The cases examined in Section 4.2 share a common structure: a translator, working sentence by sentence under time pressure, selects a target-language expression that is locally natural and semantically defensible, without fully registering that the expression carries an implicature or connotation absent from, or in tension with, the source. “Tertentu” is a perfectly ordinary Indonesian word for the context; “rumah baru” is an entirely natural way to render “a new home” outside a documentary register. The error, where it occurs, is invisible at the level of the individual sentence and becomes visible only when the sentence is evaluated against the specific pragmatic and generic demands of nature documentary narration – precisely the evaluation that the Meso and Macro tiers of ACSA are designed to make explicit and trainable.

This suggests a specific pedagogical recommendation, consistent with calls elsewhere in the audiovisual translation literature for subtitler training to move beyond technical compliance toward more explicit attention to communicative and pragmatic competence (Díaz Cintas, 2001): training that emphasises Micro-level skills (reading speed, line length, syntactic naturalness) without comparable emphasis on Meso- and Macro-level pragmatic calibration risks producing subtitlers who reliably avoid processing-effort failures while remaining vulnerable to the harder-to-detect communicative-effect failures that this study’s data show concentrate at exactly those tiers. The relative ease with which the four trainee groups in this study were able to identify and articulate Meso- and Macro-level problems, once prompted by the ACSA rubric’s explicit separation of these dimensions, suggests that the relevant pragmatic sensitivity is not absent in trainee translators so much as under-elicited by assessment frameworks that do not explicitly separate effort-side from effect-side adequacy.

A concrete training implication follows from the contrast between Group 3’s “silverback” and Group 2’s “fledging season” cases discussed in Section 4.3. Both segments involve a culturally or technically loaded English term without an established one-word Indonesian equivalent, and the two groups resolved this problem with opposite strategies – borrowing the term outright versus paraphrasing its meaning – with opposite consequences for Micro and Macro adequacy. Neither strategy is correct in the abstract; what training should cultivate is the explicit recognition that this choice exists and that it carries a predictable trade-off, so that the decision to borrow or paraphrase a given term is made

deliberately, with the ACSA framework's Micro–Macro distinction in mind, rather than by default or habit. This is, in effect, a proposal to make a specific pragmatic trade-off – one that the relevance-theoretic literature treats as a pervasive feature of constrained translation – into an explicit object of subtitle training rather than leaving it to individual translator intuition.

5.3 Limitations

Several limitations bear directly on how these findings should be interpreted. The sample of 44 scored segments, while sufficient to reveal a consistent cross-group pattern, is small relative to the full length of four documentary episodes, and the segments examined were selected by each group rather than sampled by a common, pre-specified procedure, raising the possibility of selection effects we cannot fully rule out. Scoring was conducted by the same individuals who, in most cases, also produced the written rationales justifying each score, meaning inter-rater reliability in the conventional sense – agreement between independent raters – was not assessed; the consensus-based scoring procedure, consistent with the negotiated, agreement-based view of translation quality assessment outlined in Section 3 (Lauscher, 2000), trades this form of reliability for richer, dialogically negotiated rationales. Because each episode was scored by a single group, group-level differences and episode-level differences are also fully confounded in this design and cannot be statistically disentangled. Finally, this study reports expert-assessed pragmatic adequacy, not measured viewer comprehension, engagement, or behavioural response; whether the Meso- and Macro-level adequacy patterns identified here actually predict differences in how real Indonesian-speaking viewers experience these subtitles remains an open empirical question that the present design cannot answer, though it is a natural target for the kind of viewer-reception methodology – survey, interview, controlled comparison – that relevance-theoretic subtitling research has applied in other contexts.

6. Conclusion

This article has examined the pragmatic adequacy of English-to-Indonesian subtitling in four episodes of Netflix's *Our Planet: Migrations*, using the ACSA framework's three-tier structure – Micro, Meso, Macro – to assess 44 subtitle segments scored independently by four groups of trainee translators, each responsible for one episode. Pooled results show generally high adequacy across all three dimensions, but with the Meso dimension – corresponding to the preservation of pragmatic stance and evaluative tone – showing the lowest mean and the greatest variability, and with the lowest individual scores in the dataset concentrated at points where a subtitle is locally fluent but narrows or domesticates the source narration's implicated stance or presupposed register. This pattern is consistent with a relevance-theoretic account of subtitling as an exercise in constrained pragmatic inference, in which calibrating explicature to processing constraints is more tractable than calibrating implicature and presupposition to an uncertain, culturally heterogeneous addressee.

Future work should extend this assessment design in two directions: first, by applying a common, pre-specified sampling procedure and multiple independent raters per segment, and per episode rather than per group, to strengthen the reliability of the ACSA scores themselves and to disentangle group-level

from episode-level variation; and second, by pairing expert pragmatic assessment of the kind reported here with viewer-reception data – comprehension testing, emotional-response measures, or eye-tracking of actual Indonesian-speaking audiences – to establish whether the Meso- and Macro-level adequacy patterns identified through expert scoring correspond to measurable differences in how real viewers experience and are affected by these subtitles. Such a combination would allow the ACSA framework's diagnostic granularity to be validated against, and potentially refined by, the kind of audience-facing evidence that relevance-theoretic accounts of translation ultimately need.

Declarations

Funding

This work was supported by DIPA BLU (PNBP) Universitas Mataram, Fiscal Year 2024, under an internal research grant awarded to the first author (Ketua Peneliti).

CRediT authorship contribution statement

Baharuddin: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. Lalu Ali Wardana: Investigation, Data curation, Writing – review & editing. Lalu Jaswadi Putera: Investigation, Data curation, Writing – review & editing. Desi Herayana: Investigation, Data curation, Writing – review & editing. I Gusti Agung Ayu Sasmitha Dewi Anggita Pramestya: Conceptualization (rubric design contribution), Data curation, Writing – review & editing. Alia Fitrida Camila: Conceptualization (rubric design contribution), Data curation, Writing – review & editing. Auliya Rahma: Conceptualization (rubric design contribution), Data curation, Writing – review & editing. M. Sazwan: Conceptualization (rubric design contribution), Data curation, Writing – review & editing. Duwi Wulandary: Conceptualization (rubric design contribution), Data curation, Writing – review & editing.

Acknowledgments

The authors thank to all students of the two participating subtitling classes at the English Education Program, Universitas Mataram, who carried out the subtitle scoring reported here.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author(s) used Quillbot for smoothing the language. The author(s) reviewed and edited the output as needed and take full responsibility for the content of the published article.

Clinical trial number

Clinical trial number: not applicable.

Ethics declarations

Ethics approval and consent to participate: This study did not involve clinical trials, medical interventions, or vulnerable populations. Participants were undergraduate students at Universitas Mataram, Indonesia, engaged in normal coursework activities as part of a funded internal research project (DIPA BLU/PNBP, Universitas Mataram, Fiscal Year 2024). Formal ethics committee review was not required under Universitas Mataram institutional policy for classroom-based education research of this nature. No personal, medical, or sensitive data were collected.

Consent to publish: The five named student co-authors (I Gusti Agung Ayu Sasmita Dewi Anggita Pramestya, Alia Fitrida Camila, Auliya Ramdini, M. Sazwan, and Duwi Wulandary) appear in this publication with their knowledge and consent. All seven co-authors have approved the final version of the manuscript and consent to its publication.

References

1. Baharuddin, unpublished. ACSA (Audience-Centred Subtitling Adequacy) scoring rubric: Micro, Meso, and Macro domains, 1–5 scale. Unpublished coursework material, University of Mataram.
2. Baharuddin, B., Putera, L.J., Wardana, L.A., Farmasari, S., Sukri, M., 2024. Bilingual signboards on Lombok: approaches to acquiring the translation equivalence. *World Journal of English Language* 14 (5), 612–628.
3. Baharuddin, B., Putera, L.J., Wardana, L.A., Herayana, D., Pramestya, I.G.A.A.S.D.A., Camila, A.F., Rahma, A., Sazwan, M., Wulandari, D., 2025. An in-depth translation assessment on subtitling strategies in documentary YouTube captions. *Journal of Languages and Language Teaching* 13 (3), 1149–1164.
4. Carston, R., 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Blackwell, Oxford.
5. Díaz Cintas, J., 2001. Teaching subtitling at university. in: Cunico, S. (Ed.), *Training Translators and Interpreters in the New Millennium*. University of Portsmouth, Portsmouth, pp. 29–44.
6. d'Ydewalle, G., Van Rensbergen, J., Pollet, J., 1987. Reading a message when the same message is available auditorily in another language: the case of subtitling. In: O'Regan, J.K., Levy-Schoen, A. (Eds.), *Eye Movements: From Physiology to Cognition*. Elsevier, Amsterdam, pp. 313–321.
7. Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), *Syntax and Semantics 3: Speech Acts*. Academic Press, New York, pp. 41–58.
8. Gutt, E.A., 2000. *Translation and Relevance: Cognition and Context*, second ed. St Jerome Publishing, Manchester.
9. Halliday, M.A.K., Hasan, R., 1976. *Cohesion in English*. Longman, London.
10. Kruger, J.-L., Steyn, F., 2014. Subtitles and eye tracking: reading and performance. *Reading Research Quarterly* 49 (1), 105–120.
11. Lauscher, S., 2000. Translation quality assessment: where can theory and practice meet? *The Translator* 6 (2), 149–168.

12. Levinson, S.C., 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge, MA.
13. Perego, E., Del Missier, F., Porta, M., Mosconi, M., 2010. The cognitive effectiveness of subtitle processing. *Media Psychology* 13 (3), 243–272.
14. Sperber, D., Wilson, D., 1995. *Relevance: Communication and Cognition, second ed*. Blackwell, Oxford.
15. Vermeer, H.J., 1989. Skopos and commission in translational action. in: Chesterman, A. (Ed.), *Readings in Translation Theory*. Finn Lectura, Helsinki, pp. 173–187.
16. Wilson, D., Sperber, D., 2004. Relevance theory. In: Horn, L.R., Ward, G. (Eds.), *The Handbook of Pragmatics*. Blackwell, Oxford, pp. 607–632.
17. Yusra, K., Lestari, Y.B., Nurtaat, L., Susanti, N.W.M., Hoesnie, R.K., Baharuddin, B., 2025. Address forms, social structures and social relations in postcolonial Sasak communities. *Journal of Politeness Research* 21 (2), 509–536.