Two strategies for improving machine translation from Tibetan

Nov 17, 2026 · @Drupchen

General method. Application to a specific project with its own deliverable is deliberately out of scope here.

Summary

Two strategies with different bottlenecks. **In-context** work is available today, costs nothing but prompt discipline, and is recoverable from a conversation. **Training** requires an evaluation harness that does not exist yet and a few hundred hours of expert annotation; compute and architecture are not the constraint.

The recommended order is not either-or. Run strategy one for a few weeks, measured against a harness built for strategy two, and decide about tuning only if the in-context ceiling is demonstrably hit. Nothing is wasted: the correction triples a prompt needs are the same triples a fine-tune needs, and the harness outlives any model release.

Within strategy one the leverage is lopsided. Two items carry most of it — the correction triples of §6, because the failure is a calibration problem and worked examples recalibrate where rules do not, and the sealed composition stage of §2, because the bias activates whenever source and target are co-present, which during drafting is always. The transposition table of §4 is useful and much less potent than it looks.

Revised since the first version of this advice. One recommendation is withdrawn: a blind target-only critique pass, premised on the model's judgement of the target as a text being the reliable half of its competence. It is not (§7). One is substantially rewritten: "flag what you are unsure of" produces annotation noise plus unflagged error, and is replaced by a closed list of constructions marked unconditionally (§5). Three items are new: the context question set (§1), the nominalisation-cascade repair (§4), and the channel field in the harness that stops it penalising correct conversions (§9).

Strategy one: in-context

1. The context question set

Five questions, answered before any draft or any assessment. They are listed first because omitting them is not a small inefficiency: a day's analysis can be aimed at the wrong leg of a pipeline, and the error is invisible from inside it.

Question	What it decides
Target language. Is a pivot language in use, and what for?	Whether the intermediate text should be polished or deliberately rough with decisions deferred. A pivot optimised for comprehension and a target optimised for delivery are different objects and should not be confused for one another.
Reader and register target	Which exemplars are relevant, and whether oral features are assets or noise.
Delivery conditions of the source: interpreted consecutively, dictated to a scribe, composed in writing	Which features belong to the fourth bin — present because of how the source was produced, with an addressee who may not exist in the target.
Target format: continuous prose, subtitles under cue constraints, a critical edition with apparatus	Whether channel migration applies, and what re-segmentation is licensed.
Which terminology authority governs	Whether a term decision is the model's to make or to look up.

The third and fourth are the ones usually left implicit and the ones that most change the work. Delivery conditions decide whether a segment-final particle is content or scaffolding. Target format decides whether a structural feature should be translated as text or realised in layout — and a text-to-text comparison cannot see the difference between a function correctly moved to another channel and a function lost.

Ask them once, at the top, and restate the answers in the standing instruction so they survive into later sessions.

2. The drafting pipeline

Four stages. The third carries the gain.

1. **Analyse the source.** Segmentation, connective inventory, predicate hierarchy — and then the markedness sort. The sort is not optional: without it the model attributes rhetorical intent to the grammar and invents devices to preserve. Having written *the recitation is a temporal frame and the three descriptors are co-ranked recognition criteria*, it is markedly less likely to flatten them.
2. **Enumerate the meaning units** as a flat list in the target language, with no source syntax attached. This is the representation the draft is composed from.

3. **Compose from the list**, in a context where the source is not visible. Its word order cannot then act as a template.
4. **Audit** with both texts present, for meaning-unit coverage only — not for form.

Stage 3 is the one that does work, and the reason is specific: the bias activates when source and target are co-present, and in drafting they always are. Breaking that condition deliberately is cheap and nothing else in this document substitutes for it. In practice it means a separate turn or a separate conversation, not merely an instruction to ignore the source.

Stage 4 is where discipline is easiest to lose. The audit question is *is each unit present*, not *is each element in the position its source counterpart occupied*. A rendering may be dispersed — a particle absorbed into a verb three words away, a causal link resolving two sentences later — and a position-wise check records those as losses. Before recording any absence, state what the target construction at that point is doing in its own language. If that statement cannot be made, then record the loss.

One stage this pipeline does not have, and should: a gate. A procedure that depends on the operator electing to apply it does not bind — see §10.

3. What not to preserve

A model treats every source feature as meaningful unless told otherwise, and then carries it across dutifully. This paragraph removes most of the generator of bad output and belongs in the standing instruction verbatim:

> The following are properties of Tibetan or of its production conditions, not assertions by the author. Do not preserve them, and do not treat their absence from the target as loss: one finite verb per period; verb-final position; context before predicate; repetition for durativity, iteration or intensification; concrete predication of abstract processes; flat juxtaposition linked by དེ་ནས་; ན་ as consecutive rather than hypothetical; segment-final རེད་ used to hand off a turn.

The last item is different in kind from the others and the distinction matters when something goes wrong. The first seven are **typological defaults**: absent from the target, they were never losses because they were never assertions. Segment-final རེད་ is a **function discharged outside the running text**: absent from the target, it is not a loss because its addressee — the interpreter in the room — does not exist there.

A second member of that category needs a decision rather than an instruction, because it is a conversion and not a deletion. Where the source renders its own outline in the running text (དེ་ལ་གསུམ། , then དང་པོ་ … གཉིས་པ་ …) and the target format carries structure typographically, the navigational function is genuinely needed and migrates to layout. The

working rule: render as text where the division asserts a relation and the author's terminology is doing work; convert to numbering alone where it only marks a position. Either way, tell the model which, because its default is to translate the words and leave the structure flat.

The practical consequence of the distinction is that a conversion looks exactly like a deletion in any text-to-text check, so the standing instruction has to name the conversions explicitly or the audit of §2 will flag correct work.

4. The transposition table

The model has no list of which operation to perform on meeting each source device. Supplying one helps; it helps less than it looks, because a rule gets acknowledged and then overridden by the default, where a worked example does not (§6). Use both.

The table is calibrated for French. For another Romance target it is a set of hypotheses, not results — see §11.

Source device	Required operation	Not this
Terminal `… ཟེར་མཁན་དེ་རེད་` closing a list of properties	`ce qui…, ce qui… — c'est ce qu'on appelle X`. Scan ahead for the label **before** drafting the stretch	a sequence of imperatives
`X གང་རེད་ཟེར་ན་ … Y རེད་`	keep as a question, answer introduced by a particle	a reported observation (*on verra que…*)
List totalled afterwards (`… དེ་གསུམ་དེ་ལྟན་ཅིག་ཆོགས་པ་`)	announce the count first: *Elle provient de trois choses : d'une part…*	the count after the list
Nominalised complements stacked under one copula	infinitival or relative subjects under **one** predication, present tense	a finite clause each, in the future — see below
V-གི་རེད་ generic	impersonal + present of general truth (*il arrive que*, *on*)	*vous* + simple future
Clausal repetition (iteration)	scalar adverbial: *de plus en plus*	a repeated clause
Paired reduplication (`མང་རུ་ མང་རུ་ … ཆེ་རུ་ཆེ་རུ་`)	correlative: *plus… plus…*	*de plus en plus* twice

Source device	Required operation	Not this
Lexicalised iterative (ཡང་ནས་ ཡང་དུ་)	keep: *encore et encore*	collapsing it
Verbal reduplication (durative)	adverbial or aspectual periphrasis	a repeated verb
Flat repetition of one referent across clauses	anaphoric demonstrative + one predicate	repeat the noun
Same lexeme, different referent	resolve the second to its referent	repeat it
Doublet of two distinct lexemes	render both, even where they look synonymous	merge them into one
Pro-verb དེ་འདྲ་བྱས་	short deictic tag: *Voilà ce que…*, *En procédant ainsi*, *Pour cela*	expand it back into the clause it points at
Terminal positional focus (V-final + འདི་རེད་)	cleft: *c'est là que…*	a final flat equative
Concrete figure predicated of an abstraction (ཚོས་ of a thought)	extract the content, state it analytically	transfer the figure
Explicit simile under དཔེར་ན་	transfer intact; expand for clarity if needed	abstract it away
Converb chain	gather into one complex topic, predicate once (stays left-branching)	hierarchical subordination, or a clause each
ན་	choose among five: *si* / relative / temporal (*tandis que*, *lorsque*) / causal (*étant donné que*) / manner (*en procédant ainsi*)	*si* by default
ཙམ་ limitative	*ne serait-ce que*, *simplement*, *tout juste*	drop it
ཐ་ན་	*ne fût-ce que*	drop it
བཅུག་ permissive	*laisser* + infinitive	an agentive verb
… མ་གཏོགས་ … ཡོད་མ་རེད་	a separate positive sentence	*rien d'autre*, *sans rien d'autre*

Source device	Required operation	Not this
Double negation	positive + modal adverb (*immanquablement*)	calque the two negatives
དཔེར་ན་ with no example following	drop it	*Prenons un exemple* every time
དེ་ནས་	vary, or omit	one fixed correspondent
Oral false start, self-correction, hesitation	repair into clean prose	reproduce it, or flag it

4.1 The cascade worth understanding rather than memorising

One row above deserves its mechanism, because it fires often and the drift it produces is the only one shown to change a speech act.

Where the source stacks nominalised complements under a single copula — *that is the cause of* [X], [Y], [Z] — French resists. *Le fait de bloquer la porte*, *la pacification de la souffrance*: three or four of those under *c'est la cause de* give administrative prose. So the complements are broken out as finite clauses; a finite clause needs a tense; the only tense that fits a consequence is the future; a future consequence needs a subject, and the one available is the addressee. Four steps, each locally forced, and a characterisation has become a promise. Nobody elects it.

The repair is one operation and fixes the unity, the grounding and the modality together:

> *Entendre ces mots ferme la porte des renaissances inférieures ; en pratiquer le sens apaise la souffrance du karma et des émotions perturbatrices ; en garder ne fût-ce que quatre vers libère de l'océan du samsara.*

One predication, three complements, no future, no addressee subject, and the list stays available as a premise for whatever it was grounding.

5. Flags and terminology discipline

5.1 Flag by construction, not by feeling

“Flag what you are unsure of” produces the worst available combination: annotation noise plus unflagged error, which invites trust in the unmarked text. The evidence is direct. In one real draft, of eleven `[?]` annotations four marked places a reviser changed, two marked readings that were already right, and the five most consequential errors in the file carried none — an inverted kinship relation, an iterative particle read as simultaneous, an

epistemic particle read literally three times, a personified subject supplied where the source had none, and a definition rendered as instruction. Felt lexical oddity and actual error are close to disjoint.

Replace the instruction with a closed list. Mark every instance of these regardless of confidence:

1. **A zero subject or object** whose referent had to be supplied — name the referent chosen and the alternative.
2. **A clause whose argument structure rests on a single ས-class particle** — གི་/གིས་, ཀྱི་/ཀྱིས་, གྱི་/གྱིས་, འི་/ས་, འདི་འི་/འདི་ས་ — stating which reading was taken and what the other would yield. This is a transcription-sensitivity mark, not an uncertainty mark: the model cannot repair a transcription, but it can say where the transcription is load-bearing, and that is a line resolvable from the audio in seconds.
3. **A ན་ clause** — which of the five readings was taken.
4. **A terminal ཟེར་མཁན་ or དེ་རེད་** — confirm the preceding stretch was built as a definiens and not as instruction.
5. **A doublet or a repetition** — collapsed or preserved, and why.
6. **An epistemic or modal particle** — read literally or idiomatically.
7. **A term absent from the glossary** — one rendering, used throughout, and recorded.

Drop flags for: additions required by target syntax, readings the model is in fact confident about, and anything repaired under the disfluency rule.

A flag of type 2, written out, looks like this and is worth the line it takes:

> དབྱ་འདིའི་བུ་ — read as genitive, “a child of this enemy”; if agentive, “this enemy, born as a child”, which reverses which party is the parent.

5.2 Terminology

A draft that uses two renderings of one term in adjacent paragraphs and annotates its own uncertainty is worse than one that picks and commits: the inconsistency is itself the defect, and it propagates into every downstream check.

- **One rendering per term per file.** No alternatives offered inline.
- **Position-dependent terms need the rule stated**, not left to judgement. ཞེ་སྡང་ takes one French term inside the klesha triad and another inside the ཚགས་སྡང་ pair; a model with no rule will pick one at random per occurrence.
- **Transliteration conventions belong in the glossary, not the prompt.** They recur in every file and a prompt does not carry them reliably: a draft will write *Guru*, *Shakyamuni*, *vidyadhara* where the house style is *Gourou*, *Śākyamuni*, *vidyādhara*.

- **Where a gloss is house style over a Sanskrit loan**, say so per term rather than as a principle — the same text may gloss `སྒྲ་ངན་ལས་འདས་པ་` and keep *samsara*.

6. What goes in the prompt

6.1 Correction triples — the highest-leverage item

Five to ten worked corrections will move output further than any quantity of rules. The reason is mechanical: the failure is a **calibration** problem, not a knowledge gap. The model already writes the target language well; what it lacks is which target text to produce here, and a prior is moved by examples, not by instructions it can acknowledge and then override.

Keep them as **triples**: source segment, draft, correction. The triple teaches the operation; the pair teaches only the output, which generalises worse.

Select for patterns, not facts. One triple per pattern beats five of the same pattern. The patterns worth a slot each, in order: a definition rendered as instruction; a repetition reproduced where it should scalarise; a `མ་གཏོགས་` clause calqued; a zero subject resolved wrongly; a list whose count should have been announced first; a pro-verb expanded into a restatement.

Validity condition, and it is not optional. The source must be unchanged between the draft and the correction. A pair whose source was revised in between encodes a phantom rule: it credits the draft with an error it did not make, or penalises a reading that was correct when made. One real instance — an apparent comprehension failure, an inverted kinship relation, traced to a genitive particle standing in the transcription where the audio had an agentive. The draft had read its input correctly.

In practice: stamp every pair with the transcription revision it was made against, and discard pairs that straddle a revision. For a project where transcriptions continue to be corrected after drafts are made, this is a standing hygiene requirement rather than a one-off check.

If a decision or precedent store can hold these retrievably, that is where they belong, so a new draft pulls the nearest few rather than depending on what happens to be in context.

6.2 Register exemplars — and why they must not be translations

“Good French” is too weak an instruction, because the model’s prior for *French translated from Tibetan* beats its prior for *French*. That corpus is small and dominated by one tradition, with a direct consequence: **a competent discourse-level translation resembles the prior less closely than a poor literal one does.**

So name the target genre independently of translation — explanatory prose of a given authority and formality — and supply two or three exemplars of **originals** in the target language. Translated exemplars re-import source-shaped syntax and defeat the purpose.

Note the tension with §6.1, because it is real and both sides are needed. Aligned source-target pairs define fidelity. Non-translated originals define the target. They are different corpora and neither substitutes for the other: a prompt with only pairs produces faithful-sounding calque, and a prompt with only originals produces fluent drift.

7. The limit

This section supersedes an earlier recommendation in this series — a blind target-only critique pass, on the reasoning that a model's competence in the target language rests on a vast corpus of native prose while its sense of that language *as a translation target* rests on a tiny one, so judgement of the target as a text should be the reliable half. That reasoning is wrong, and the test was direct.

Asked to critique an expert translator's French against the Tibetan, a model produced six findings. Four were withdrawn under correction. Each died the same way: an element of the source was sought at the position its correspondent should occupy, not found there, and recorded as a loss — while the correspondent stood elsewhere in the sentence. One of the withdrawals asserted a constraint on French topicalisation that does not exist. The six had already accumulated into a pattern, which collapsed with them.

The two that survived were both **countable**. So was every residue of the four that died.

Trust the model for	Do not trust it for
a term rendered two ways in one file	whether a construction sounds native
a lexeme repeated across adjacent sentences	whether a construction is licensed
markers stacked at one joint	whether a sentence is too heavy
agreement, tense and person switches	which register a text is in
a predication count that does not match the source's	whether a turn of phrase is idiomatic
a meaning unit absent from the target	whether the prose is any good

The left column is mechanical and checkable by the person receiving it in seconds. The right column requires a judgement the model oscillates on — and oscillates, not errs consistently, which is worse: the same construction was read first as a deliberate spoken-

register norm and then, after a remark about register, as a defect. A bias can be corrected for. An oscillation tracks whatever was last said in the conversation.

Two working consequences. Ask for countable checks explicitly and by name, rather than for a critique. And treat any register verdict the model volunteers as an unprompted opinion rather than a finding, including the confident ones — especially the confident ones, since §11.2 of the companion analysis explains why granular, individually checkable findings accumulate into unearned confidence.

Strategy two: trained

8. The data

8.1 The obvious corpus is poison as positive data

Supervised fine-tuning on published Tibetan-to-Romance translation **reinforces** the bias it is meant to remove. The corpus is small and dominated by the philologically literal tradition, so frequency in it measures the prevalence of a convention, not its fidelity. Scraping more makes the problem worse, and does so invisibly, because the resulting model's outputs will match the corpus better at every step.

The same corpus is valuable **inverted**. Because the failure is a pull toward one specific wrong style rather than diffuse incompetence, labelled negatives are unusually informative here — which reverses their normal low value. A set of literal renderings marked as what not to produce carries real signal.

8.2 The shape that works: correction triples

What carries signal is the operation, not the output: `(source, draft, corrected draft)`. A few hundred well-curated triples would plausibly outperform tens of thousands of published pairs, for the same reason §6.1 gives — the model's gap is calibration, and preference-style training on corrections targets calibration directly. A LoRA is the right weight class; this is not a capability that needs full fine-tuning.

The triples are generated as a by-product of ordinary work. Anyone reformulating machine drafts is producing them already; the only additional cost is capturing them in a stable form.

8.3 Curation is prior to, and more load-bearing than, the model

Extraction inherits whatever the exemplar set contains. Selecting for **discourse-level transposition** rather than philological transparency is a selection criterion applied by hand, not something a downstream layer can correct for — the layer has no independent signal against which to recognise that the set is skewed.

This is where the real cost sits. Telling the two apart requires linguistic training, native command of the target, fluency in the spoken source and formation in its exegetical tradition, held simultaneously. That combination is rare enough that most of the relevant knowledge is unwritten, which means the gap in any model's training reflects a gap in the literature rather than an accident of sampling. More corpus does not fix it.

8.4 One hygiene requirement carried over

The validity condition of §6.1 applies to training pairs with more force than to prompt pairs, because a phantom rule in a prompt can be noticed and a phantom rule in weights cannot. Every triple needs the source revision it was made against recorded, and pairs straddling a revision discarded. For a project where transcriptions are corrected on a rolling basis, this is the difference between a usable set and a set that teaches the model to repair errors that were never made.

9. The evaluation harness

Build this first, whether or not any tuning happens. Three reasons, in order of force.

Reference-overlap metrics — BLEU, chrF, COMET-style scoring — measure surface similarity to a reference translation. In this language pair that means they **actively select for literalism**, so tuning against them optimises the exact bias the project exists to remove. There is no version of this work that can use them as its objective.

The harness is model-independent and outlives any release, which the prompt engineering of §§1–7 does not.

And it works at inference time as a self-check, which is the only mechanism identified anywhere in this document that makes a procedure **gate** output rather than advise it (§10).

9.1 Three components

1. **Meaning-unit coverage.** Per source segment, an expert-authored list of the units it carries. Score each candidate per unit: delivered / shifted / thinned / absent. This is the only component that requires expert authorship, and it is the component that does the real work.

2. **Countable battery**, fully automatable, and matching exactly the left column of §7: term consistency within a file; lexeme repetition across adjacent sentences; marker stacking at one joint; predication count against the source; person and tense switches; glossary and transliteration conformance.
3. **Register**, by **blind pairwise preference** between candidates, not by rubric. Rubric scoring of register is precisely what §7 shows to be unreliable, and a rubric gives the unreliability a number, which is worse than no score.

9.2 The design constraint that is easy to miss

A naive coverage metric penalises correct work, and does so silently.

Two classes of source feature are properly **not** rendered as target text: those whose addressee does not exist in the target, and those whose function migrates to another channel — a sapche realised as a heading, a segment-final handoff particle correctly dropped. A unit list that records only present / absent marks both as absence and scores the right answer down.

So the unit list needs a per-unit **channel** field, with at least three values: *carried in text*, *carried outside the text* (layout, structure, apparatus), *correctly absent*. Without it, every conversion the harness meets is a false positive, and the harness will train or select against exactly the operations that distinguish good work here.

9.3 The colophon test bed

The best available corpus for building and calibrating the harness, better than continuous teaching, for four reasons: colophons are formulaic, so the slots are fixed and scoring is mechanical; they are numerous; they are widely published with the source facing; and nobody can claim their structure was a stylistic choice.

A colophon is one head-final period with a conventional slot order and the matrix verb last: occasion or request, the requester's honorific phrase, place, date, the author's self-deprecating epithet, aspiration, then the verb with its agent marked late. The slots are ordered by convention, not by logic — which is why rendering them in source order produces a target sentence whose predicate arrives sixty words in or not at all, and why the translator then has to supply connectives to hold the fragments together.

Four mechanical measures, scorable without expert judgement:

Measure	Scoring
Matrix verb placement	early / stranded / lost

Measure	Scoring
Supplied connectives	count of *thus*, *therefore*, *hence* and equivalents that no source particle licenses
Author's epithet	rendered as a slot / promoted to a predication
Slot order	source order preserved / re-emitted in target order

Run it on twenty before committing to two hundred, to check the rubric catches what it should. If the rate of invented connectives is as high as practitioners report, that is also the strongest available argument for the whole procedure, on the one genre where the structure cannot be defended as a choice.

10. Sequencing, cost, and the ceiling

10.1 Order

1. Answer the question set (§1). One question would have redirected a day's work.
2. Run the in-context pipeline (§§2–6), accumulating correction triples as a by-product.
3. Build the harness (§9), seeded on colophons.
4. Measure. Tune only if the in-context ceiling is demonstrably hit.

Nothing is wasted whichever way step 4 goes: the triples a prompt needs are the triples a tune needs, and the harness is required either way.

10.2 Cost

Compute and architecture are not the constraint. The constraint is expert annotation: on a rough estimate, a few hundred hours to produce a usable triple set and a seeded harness. That is the part nobody else can do, and it is worth stating plainly before the project is scoped, because the engineering looks like the hard part and is not.

10.3 A procedure must gate, not advise

The most uncomfortable finding in this series, and the one with the clearest design consequence. A procedure that depends on the operator electing to apply it does not bind. The four-step assessment procedure was written down, and then bypassed twice in one working session by the party who had written it — both times while moving quickly and feeling well evidenced.

So the checks have to be mechanical and interposed, not advisory and adjacent. In practice that means the countable battery of §9.1 runs on every draft automatically rather than when someone remembers; the flag list of §5.1 is a required output format rather

than a request; and the audit of §2 stage 4 asks for the per-unit channel field (§9.2) to be filled in, because a form with a blank in it gets noticed where a skipped instruction does not.

10.4 What neither strategy buys

Genuine indeterminacy does not resolve with better prompting or more weights. A bare past stem that could be narrative or directive; an anaphor pointing outside the segment; an agentive-genitive ambiguity in the transcription; a crux where the sense turns on an idiom in the figure's own domain rather than on the text. These need context the model does not have and judgement it cannot acquire from text alone.

The reachable ceiling is therefore **a good first draft with its decisions marked** — not a finished translation. Which is why §5.1 matters more than it looks: if the ceiling is a flagged draft, the quality of the flagging is the quality of the product.

Both

11. Conditions on both strategies

Three, each of which invalidates work silently if left unmet.

Source stability. A draft/correction pair is evidence only if the source did not change between the two. Record the revision each pair was made against; discard pairs that straddle a revision. In a project where transcriptions are corrected on a rolling basis this is continuous hygiene, not a one-time check (§6.1, §8.4).

Rules are per language pair. The transposition table of §4 is derived for Tibetan–French. For another Romance target it is a set of hypotheses. Three differences are large enough to change the results:

- **Null subjects.** Spanish, Italian and Portuguese permit unexpressed subjects; French does not. Since the source drops arguments freely, a French translator must resolve a zero subject at every occurrence and can resolve it wrongly — a null-subject target can sometimes carry the underspecification forward. The error class that flag type 1 exists to catch is largely a French error class.
- **Nominalisation weight.** The cascade of §4.1 begins in French's resistance to stacked nominalised complements. The other three have more productive infinitival-subject constructions, and Portuguese has the inflected infinitive, which French lacks. Step one may not fire, and with it neither the future nor the addressee subject.

- **Topicalisation tolerance.** Left dislocation with clitic resumption is more freely available in the written registers of Italian, Spanish and Portuguese than in written French. Advice calibrated for French is pitched at the most restrictive member of the family.

Untested and not to be assumed: whether the markedness of repetition and the abstraction differential hold at the same strength across the family. Both are stated here on French evidence.

Source-side findings do travel. The division is not between languages but between sides of the operation. Everything that concerns comprehension — the definition pattern and its question-answer variant, the range of ཀ་, zero-argument resolution, the epistemic particles, doublets against repetitions, the pro-verb, the particle-series ambiguity, and the fourth bin — arises before a target language is involved and goes to any target unchanged. It is only the encoding side that needs re-deriving. A set of transposition rules is pair-specific; a set of comprehension failures is not.