

**Human Evaluation Table for Machine Translation**

Set	Index	Source (EN)	Corrected Reference (NSO)	MT Output (NSO)	Evaluator Comments
Dev	307	He died in Osaka on Tuesday.	O hlokofaletše Osaka ka Labobedi.	O hlokofetše Osaka ka Labobedi.	Perfectly fluent and semantically aligned with the reference. The original reference is more accurate by using “hlokofaletše” which makes it clearer that Osaka is a place.
Dev	859	Some countries have exceedingly draconian punishments even for first time offenses; these can include prison sentences of over 10 years or death.	Dinaga tše dingwe di na le dikotlo tše šorošoro le go ditshenyo tša mathomo; tše di ka akaretša dikatlolo tša go feta mengwaga ye 10 tša go ya kgolegong goba lehu.	Dinaga tše dingwe di na le dikotlo tše šoro kudu gaešita le bakeng sa melato ya mathomo; tše di ka akaretša dikotlo tša kgolego tša ka godimo ga nywaga e 10 goba lehu.	Good adequacy, the model output is slightly less formal than the corrected reference. Which is more aligned with everyday human spoken words, such as “dikotlo tše šoro”
Dev	845	However, nowadays there are established graffiti artists, graffiti events and "legal" walls. Graffiti paintings in this context often resemble artworks rather than illegible tags.	Le ge go le bjalo, matšatši a go na le borabokgabo ba dithalwa bao ba tšweletšego, meletlo ya dithalwa le maboto a go ba "molaong". Dithalwa tša go pentwa ka se sebaka gantši di emetše mošomo wa bo kgabo go na le ditag tša go se be molaong.	Lega go le bjalo, matšatšing a lehono go na le bathadi ba diswantšho tša diswantšho tša diswantšho tša diswantšho tša diswantšho tša diswantšho tša ... ... ... diswantšho tša.	Severe hallucination and repetition. Output is completely unusable.
Dev	552	Maintaining a healthy energy balance, practicing effective hydration habits, and understanding the various aspects of supplementation practices can help athletes	Go boloka tekanyo yeo e phelegilego ya enetši, go ithuta mekgwa ye mekaone ya go dula o na le meetse, le go kwešiša dikarolo tša go fapana tša mekgwa ya go emela ye mengwe go ka thuša boramabelo go kaonafatša mešomo ya bona le go oketša go	Go boloka teka-tekano e botse ya matla, go itlwaetša mekgwa e atlegago ya go nwa meetse le go kwešiša dibopego tše di fapa-fapanego tša mekgwa ya go nwa dijo tše di nago le phepo go ka thuša bo-ramabelo go kaonefatša bokgoni bja bona le go oketša	Both are good, but corrected reference provides a more precise and faithful translation of the original sentence, particularly in its treatment of supplementation practices, making it the better choice for accuracy. The model output, although well-constructed and clear, interprets supplementation more generally as nutritious

		improve their performance and increase their enjoyment of the sport.	ipshina ga bona ka papadi.	lethabo la bona la go bapala dipapadi.	food consumption, which slightly diverges from the original focus. Therefore, for a translation that closely mirrors the original meaning, corrected reference is preferable.e.
Dev	144	That means everything has to be put back," said researcher at the Rift Valley Research Service in Ethiopia and a co-author of the study, Berhane Asfaw.	Seo se ra gore se sengwe le se sengwe se swanetše go boetšwa morago," gwa bolela monyakišiši go la Rift Valley Research Service ka Ethiopia le mongwadi-thuši, Berhane Asfaw.	Seo se bolela gore selo se sengwe le se sengwe se swanetše go bušetšwa morago", go boletše monyakišiši wa Rift Valley Research Service kua Ethiopia le mongwadi-gotee wa nyakišišo, Berhane Asfaw.	Fluent and semantically faithful. Small stylistic variation only but the model output is much better, fluent and mirrors the tone of the english sentences, mainly in the words " bušetšwa morago" which closely aligns with structure in the original sentence.
DevTest	589	These couples may choose to make an adoption plan for their baby.	Baratani ba ba ka kgetha go dira lenaneo la go fana ka ngwana wa bona.	Banyalani ba ba ka kgetha go dira leano la go godiša ngwana wa bona.	Major variation in word choice ('godiša' vs. 'fana ka'), meaning "raise" and "give" - the latter aligns with adoption more while the former takes a stance that the couple will raise their own child. Also the model output uses the word "Banyalani" which means a married couple instead of "Baratani" which means "lovers", which is more semantically aligned with "couples".
DevTest	269	The fatalities were Nicholas Alden, 25, and Zachary Cuddeback, 21. Cuddeback had been the driver.	Dikgobadi e be e le Nicholas Alden, 25, le Zachary Cuddeback, 21. Cuddeback e be e le mootledi.	Bahlaselwa e be e le Nicholas Alden wa nywaga e 25 le Zachary Cuddeback wa nywaga e 21. Cuddeback e be e le mootledi.	Different term for 'victims' vs. 'injured'. Both are acceptable, corrected references are slightly more accurate and do not break grammar as seen in MT output by using the word "ngwaga" (year) instead of "mengwana" (age).



		than twenty other people.		ka godimo ga ba masomepedi.	
--	--	------------------------------	--	--------------------------------	--

### Summary and Explainability:

In multilingual machine translation, especially for low-resource languages like Northern Sotho (Sepedi), models often struggle with fluency, adequacy, and semantic alignment. One of the most critical issues is hallucination — where the model generates repetitive or nonsensical output due to uncertainty or missing vocabulary. A clear example is the sentence about USA Gymnastics, where the model outputs “dipapadi tša go bapala” (“games for playing”) in a loop instead of translating “gymnastics,” showing it lacked a confident match and defaulted to familiar fragments. A similar error occurs in the graffiti example, where the model produces “diswantšho tša diswantšho tša...” repeatedly, making the output completely unusable. These are caused by decoder instability, especially when translating abstract or unfamiliar domain content.

Other examples highlight semantic drift or partial adequacy. For instance, in a health-related sentence about “supplementation practices,” the model interprets it more broadly as nutritious food, diverging slightly from the intended meaning. In the adoption sentence, it uses “godiša” (raise) instead of “fana ka” (give up), which shifts the intent from adoption to parenting. Another subtle example involves translating “victims” as “dikgobadi” (injured), which isn’t entirely incorrect but lacks the precision of “bahlaselwa”.

Even when translations are mostly fluent, they can still lose tone or intent. For example, the Holocaust sentence repeats “tlhompho” (respect) unnecessarily, while the car bomb sentence uses slightly simplified phrasing. These kinds of inconsistencies reflect both the limited representation of Sepedi in training data and the challenges of handling formal, emotional, or abstract English in an under-resourced target language. Together, these issues reveal that while models like NLLB are powerful, they still require careful tuning, data augmentation, and human post-editing to ensure high-quality outputs in low-resource contexts.