A List of Topics

topic	proportion (%)	
	ne–en	si–en
General	18.3	24.1
History	6.5	15.1
Science	7.4	12.7
Religion	8.9	10.5
Social Sciences	10.2	6.9
Biology	6.3	9.1
Geography	10.6	4.6
Art/Culture	6.7	8.3
Sports	5.8	6.7
Politics	8.1	N/A
People	7.4	N/A
Law	3.9	2.0

Table 1: Distribution of the topics of the sentences in the dev, devtest and test sets according to the Wikipedia document they were sampled from.

B Statistics of automatic filtering and manual filtering

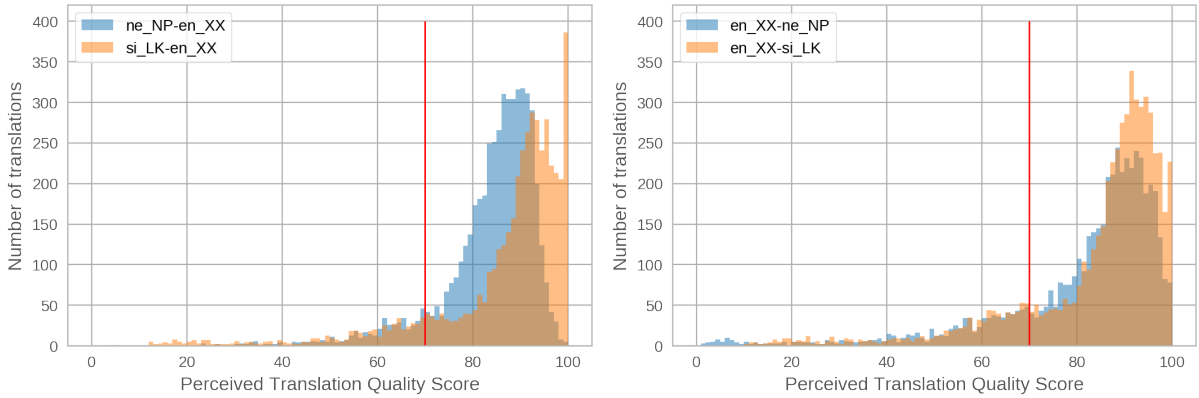

Figure 1: Histogram of averaged translation quality score. We ask three different raters to rate each sentence from 0–100 according to the perceived translation quality. In our guidelines, the 0–10 range represents a translation that is completely incorrect and inaccurate; the 11–29 range represents a translation with few correct keywords, but the overall meaning is different from the source; the 30–50 range represents a translation that contains translated fragments of the source string, with major mistakes; the 51–69 range represents a translation which is understandable and conveys the overall meaning of source string but contains typos or grammatical errors; the 70–90 range represents a translation that closely preserves the semantics of the source sentence; and the 90–100 range represents a *perfect* translation. Translations with averaged translation score less than 70 (red line) are removed from the dataset.

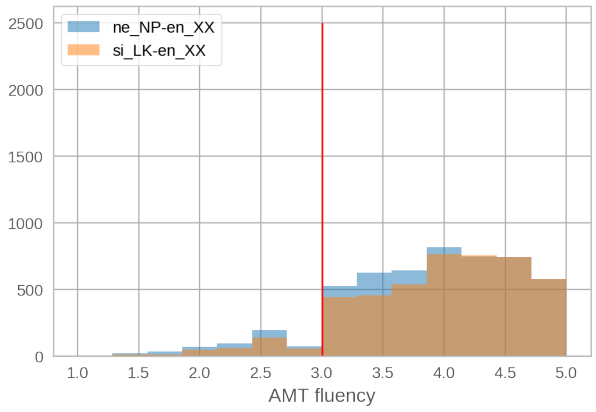

Figure 2: Histogram of averaged AMT fluency score of English translations. We ask five different raters to rate each sentence from 1–5 according to its fluency. In our guidelines, the 1–2 range represents a sentence that is not fluent, 3 is neutral, while the 4–5 range is for fluent sentences that raters can easily understand. Translations with averaged fluency score less than 3 (red line) are removed from the dataset.

	Nepali–English	English–Nepali	Sinhala–English	English–Sinhala
Automatic filtering	14%	18%	24%	7%
Manual filtering				
Translation quality	10%	19%	13%	16%
Fluency	10%	-	17%	-

Table 2: Percentage of translations that did not pass the automatic and manual filtering checks. We first use automatic methods to filter out poor translations and send those translations back for rework. We then collect translations that pass the automatic filtering and send them to two human quality checks, one for adequacy and the other for fluency. Note that the percentage of sentences that did not pass manual filtering is among those sentences that passed the automatic filtering.

C List of Wikpedia Documents

domain	document/gloss	topic
en.wikipedia.org	Astronomy	Science
en.wikipedia.org	History of radar	History
en.wikipedia.org	Shoe	General
en.wikipedia.org	Tire	General
en.wikipedia.org	Indian cuisine	Art/Culture
en.wikipedia.org	IPhone	General
en.wikipedia.org	Apollo program	History
en.wikipedia.org	Chess	General
en.wikipedia.org	Honey	General
en.wikipedia.org	Police	Law
en.wikipedia.org	Desert	Geography
en.wikipedia.org	Slavery	Social Sciences
en.wikipedia.org	Riddler	Art/Culture
en.wikipedia.org	Diving	Sports
en.wikipedia.org	Cat	Biology
en.wikipedia.org	Boxing	Sports
en.wikipedia.org	White wine	General
en.wikipedia.org	Creativity	Social Sciences
en.wikipedia.org	Capitalism	Social Sciences
en.wikipedia.org	Alaska	Geography
en.wikipedia.org	Museum	General
en.wikipedia.org	Lifeguard	General
en.wikipedia.org	Tennis	Sports
en.wikipedia.org	Writer	General
en.wikipedia.org	Anatomy	Science
si.wikipedia.org	Qoran	Religion
si.wikipedia.org	Dhammas	Religion
si.wikipedia.org	Vegetation	Science
si.wikipedia.org	Names of Colombo Students	History
si.wikipedia.org	Titanic	History
si.wikipedia.org	The Heart	Biology
si.wikipedia.org	The Ear	Biology
si.wikipedia.org	Theravada	Religion
si.wikipedia.org	WuZetian	History
si.wikipedia.org	Psychoanalisis	Science
si.wikipedia.org	Angulimala	Religion
si.wikipedia.org	Insurance	General
si.wikipedia.org	Leafart	Art/Culture
si.wikipedia.org	Communication Science	Science
si.wikipedia.org	Pharaoh Neferneferuaten	History
ne.wikipedia.org	Nelson Mandela	People
ne.wikipedia.org	Parliament of India	Politics
ne.wikipedia.org	Kailali and Kanchanpur	Geography
ne.wikipedia.org	Bhuwan Pokhari	Geography
ne.wikipedia.org	COPD	Biology
ne.wikipedia.org	KaalSarp Yoga	Religion
ne.wikipedia.org	Research Methodology in Economics	Social Sciences
ne.wikipedia.org	Essay	Social Sciences
ne.wikipedia.org	Mutation	Science
ne.wikipedia.org	Maoist Constituent Assembly	Politics
ne.wikipedia.org	Patna	Geography
ne.wikipedia.org	Federal rule systen	Law
ne.wikipedia.org	Newari Community	Art/Culture
ne.wikipedia.org	Raka's Dynasty	History
ne.wikipedia.org	Rice	Biology
ne.wikipedia.org	Breastfeeding	Biology
ne.wikipedia.org	Earthquake	Science
ne.wikipedia.org	Motiram Bhatta	People
ne.wikipedia.org	Novel Magazine	Art/Culture
ne.wikipedia.org	Vladimir Putin	Politics
ne.wikipedia.org	History of Nelali Literature	History
ne.wikipedia.org	Income tax	Law
ne.wikipedia.org	Ravi Prasjal$^+$	People
ne.wikipedia.org	Yogchudamani Upanishads$^+$	Religion
ne.wikipedia.org	Sedai$^+$	Religion

Table 3: List of documents by Wikipedia domain, their document name or English translation, and corresponding topics. The document name has an hyper-reference to the original document. $^+$ denotes a page that has been removed or no longer available at the time of this submission.

D Examples from *devtest*

En→Ne		
Source		It has automatic spell checking and correction, predictive word capabilities, and a dynamic dictionary that learns new words.
References	**A**	यसमा स्वचालित हिज्जे जाँच र सुधार छ , भविष्यवाणी शब्द क्षमताहरू , र गतिशील शब्दकोश हुन्छ जसले नयाँ शब्दहरू सिक्छ ।
	B	यसमा स्वचालित हिज्जे जाच्ने तथा सच्याउने , शब्दहरूको अनुमान गर्ने , तथा नयाँ शब्दहरू सिक्ने स्फुर्त शब्दकोश हुन्छ ।
System		यसमा स्वचालित हिज्जे जाँच र सुधार , पूर्वानुमान शब्द क्षमता र नयाँ शब्द सिक्ने गतिशील शब्दकोश छ ।
Source		The academic research tended toward the improvement of basic technologies, rather than their specific applications.
References	**A**	शैक्षिक अनुसन्धानले उनीहरूको विशिष्ट अनुप्रयोगहरूको सट्टा आधारभूत प्रविधिको सुधारको पक्षमा जोड दिए ।
	B	यो शैक्षणिक अनुसन्धान सामान्य प्रविधिको सुधार तर्फ ढलकिएको छ , नाकि तिनिहरूको विशेश प्रयोग तर्फ ।
System		प्राध्यापक अनुसन्धानले उनीहरूको विशिष्ट अनुप्रयोगमन्दा पनि आधारभूत प्रविधिको सुधारतिर टेवा पुर्यायो ।
Ne→En		
Source		पुरानो समयमा राजालाई सल्लाह दिने सभा ' संसद ' कहलाउँथ्यो ।
References	**A**	In the past, the assembly that advised the king were called 'parliament'.
	B	In old times the counsil that gave advice to the king was called 'parliament'.
System		In old times the council of counsel to the king was 'Senate'.
Source		कार्यकर्ताका रुपमा अफ्रिकन नेशनल कंग्रेसमा आबद्ध भए ।
References	**A**	As a worker African Mandela joined the Congress party.
	B	He joined the African National Congress as a activist.
System		As a worker, he joined the African National Congress.
En→Si		
Source		Iphone users can and do access the internet frequently, and in a variety of places.
References	**A**	අයිෆෝන් භාවිත කරන්නන්ට නිතරම සහ විවිධ ස්ථානවලදී අන්තර්ජාලයට පිවිසිය හැකිය .
	B	Iphone පරිශීලකයින් හට නිතරම විවිධ ස්ථානවලදී අන්තර්ජාලය වෙත පිවිසීමට හැකිය .
System		අයිෆෝන් භාවිතා කරන්නන්ට පුළුවන් වගේම අන්තර්ජාලයට නිතර පිවිසෙන්නත් පුළුවන් . ඒ වගේම විවිධ තැන්වල
Source		In Serious meets, the absolute score is somewhat meaningless.
References	**A**	සැබෑ තරඟ වලදී ලකුණු සැසඳීම තේරුමක් නැති ක්රියාවකි .
	B	වැදගත් උළෙල වල , නිරපේක්ෂ ලකුණු තරමක් නිෂ්ඵල ය .
System		සීරියස් හමුවේ , නිරපේක්ෂ ලකුණු කිසියම් තේරුමක් නැති
Si→En		
Source		තර්ජන , ශාරීරික හිංසනය , දේපල හානිය , පහර දීම සහ මරාදැමීම මෙම දඬුවම්ය .
References	**A**	Threatening, physical violence, property damage, assault and execution are these punishments.
	B	Threats, bodily violence, property damages, assaults and killing are these punishments.
System		Threats, physical harassment, property damage, strike and killing this punishment.
Source		අධ්‍ය යාපනයෙන් පසු හෝ පවුලේ යුතුකම් ඉටු කරන්නට හෝ රෝග තත්වයන් නිසා සංඝයා උපසම්පදාවෙන් නිතරම ඉවත් වෙති .
References	**A**	After education priests leave ordination in order to fulfill duties to the family or due to sickness.
	B	Sangha is often abandoned because of education or after fulfilling family responsibilities or because of illness.
System		After education or to fulfill the family's disease or disease conditions, the companion is often removed from substance.

Table 4: Examples of sentences from the En-Ne, Ne-En, En-Si and Si-En *devtest* set. System hypotheses (System) are generated using the semi-supervised model described in the main paper using beam search decoding.