Synopsis of the generated datasets

- *final_linked_set_v4.xls*

This is generated using the [Hindi word, Eng. concept pair], [आ_1, come_1] which were both found in the same row in our initial hin-eng synset linked data.

Column definitions:

index: index of the row (not necessarily maintains contiguity). Is star marked if the hindi word: concept word pair has previously been seen in the data set or not.

hin: The Hindi word(आ) along with its occurrence number (in the mrs concept dict, 1) separated by an underscore

eng: the English concept word**(**come) along with its occurrence number(in the mrs concept dict, 1) separated by an underscore

hin_id: the id of this word's synset entry in the hindi synset data

eng_id: the eng synset id of this word's synset entry in the hin:english linked synset data

hin_gloss: meaning of the hindi word in hindi

eng_gloss: list of English meanings of the eng concept word

hin_ex: usage example of the word/concept in hindi.

eng_ex: usage example of the word/concept in English.

Duplicate: flag signifying whether this particular hindi word, eng concept pair is present anywhere else before this row or not.

index	hin	eng	hin_id	eng_id	hin_gloss	eng_gloss
0	आ_1	come_1	30934	2729963	किसी के अंतर्गत होना	['come under, be classified or included']
1**	आ_1	come_1	11882	341917	किसी भाव या अवस्था आदि का उत्पन्न होना	['come to pass', ' arrive, as in due course']

hin_ex	eng_ex	Duplicate
"बनारस उत्तर-प्रदेश में आता है । / यह कथा रामायण में आती है ।"	['fall into a category', 'This comes under a new heading']	No
"आज के हास्य कवि सम्मेलन में बहुत आनंद आया ।"	['The first success came three days later', 'It came as a shock', 'Dawn comes early in June']	Yes

- *final_coarse_linked_set_v2.xls*

Some hindi words and English concept pairs were not found in the same row, but these hindi words were found individually in the linked synset data with some other English synsets (none of which contained the original eng concept word we were looking for).

Column definitions:

index: index of the row (not necessarily maintains contiguity). Is star marked if the hindi word: concept word pair has previously been seen in the data set or not.

hin: The Hindi word(अभिवादन) along with its occurrence number (in the mrs concept dict, 1 here) separated by an underscore

eng_conc: The English concept word**(**greeting) along with its occurrence number (in the mrs concept dict, 2 here) separated by an underscore. ***Note: this was not found in the linked data set in the same row.***

eng_syns: the corresponding synset of English words which was found along with the hin word in the same row in the linked dataset.

hin_id: the id of this word's synset entry in the hindi synset data

eng_id: the eng synset id of this word's synset entry in the hin: English linked synset data

hin_gloss: meaning of the hindi word in hindi

eng_gloss: list of English meanings of the eng concept word

hin_ex: usage example of the word/concept in hindi.

eng_ex: usage example of the word/concept in English.

Duplicate: flag signifying whether this particular hindi word, eng concept pair is present anywhere else before this row or not.

index	hin	eng_conc	eng_syns	hin_id
0	अभिवादन_1	greeting_2	['salute', 'salutation']	3006
1*	अभिवादन_1	greeting_2	['praise', 'congratulations', 'kudos', 'extolment']	3632

eng_id	hin_gloss	eng_gloss	hin_ex	eng_ex	Duplicate
6688913	किसी के प्रति आदर भाव दिखाने की क्रिया	['an act of honor or courteous recognition']	"मुख्य अतिथि के अभिवादन के बाद कार्यक्रम शुरू हुआ ।"	['a musical salute to the composer on his birthday']	No
6693198	किसी वस्तु, व्यक्ति, आदि या उनके गुणों या अच्छी बातों के संबंध में कही हुई आदरसूचक बात	['an expression of approval and commendation']	"प्रशंसा से सभी खुश और प्रोत्साहित होते हैं ।"	['he always appreciated praise for his work']	Yes

- *hindi_nf_set_v2.xls*

Column definitions:

Some hindi words are not found in any of the hindi synset list in any of the rows in the linked hn-eng synset data. So the hindi-only synset data was looked into for these hindi words and any occurrences were added to this dataset.

index: index of the row (not necessarily maintains contiguity). Is star marked if the hindi word: concept word pair has previously been seen in the data set or not.

hin_wx: The Hindi word in wx format (AbakArI_1) along with its occurrence number (in the mrs concept dict, 1 here) separated by an underscore

hin_utf: The Hindi word in utf format(आबकारी_1) along with its occurrence number (in the mrs concept dict, 1 here) separated by an underscore

conc_eng: The English concept word(greeting) along with its occurrence number (in the mrs concept dict, 2 here) separated by an underscore.

conc_hn: the hindi word's concept in Hindi as found in the hindi synset data.

hn_ex: the usage example of the hindi word/concept in Hindi.

Duplicate: flag signifying whether this particular hindi word, eng concept pair is present anywhere else before this row or not.

index	hn_wx	hn_utf	conc_eng
0	AbakArI_1	आबकारी_1	excise_1
1*	AbakArI_1	आबकारी_1	excise_1

conc_hn	hn_ex	Duplicate
शराब बनने और बिकने की जगह	"रामू शराबघर में काम करता है ।"	No
मादक वस्तुओं से संबंध रखनेवाला एक सरकारी विभाग	"महेश आबकारी में काम करता है । / आज आबकारी मंत्री आनेवाले हैं ।"	Yes

- *augmented_mrs_v3.xls*

This was generated using the [Hindi word, Eng. concept pair] which were found in the same row in the linked data. So if the word, meaning pair अब_1, now_3 was found then all the hindi synonyms of अब: अभी, फ़िलहाल, फिलहाल,इस समय,इस वक्त ,....... were taken from the hindi synset list present in the row, along with the eng_conc(now_3) and were used to construct a novel hindiword: concept dictionary as an augmentation to the initial concept dict.

Column definitions:

conc_label_wx: the hindi concept label in wx format

conc_label_utf: the hindi concept label in utf format

conc_in_eng: The English concept as found for the original word

mrs_conc: the mrs concept as found for the original word.

conc_label_wx	conc_label_utf	conc_in_eng	mrs_conc
aBI_1	अभी_1	now_3	_now_a_1
PZilahAla_1	फ़िलहाल_1	now_3	_now_a_1
PilahAla_1	फिलहाल_1	now_3	_now_a_1
isa+samaya_1	इस+समय_1	now_3	_now_a_1

isa+vakwa_1	इस+वक्त_1	now_3	_now_a_1
aWa_1	अथ_1	now_3	_now_a_1
samprawi_1	सम्प्रति_1	now_3	_now_a_1
aBU_1	अभू_1	now_3	_now_a_1
aBE_1	अभै_1	now_3	_now_a_1
saxya_1	सद्य_1	now_3	_now_a_1