Building Neural Net Machine Translation Systems Using Interlinear Glossed Texts

(1) a. Training Data:

        Tha   a   athair nas   sine   na   a   mhàthair.
be.pres 3sm.poss father comp old.cmpr comp 3sm.poss mother

‘His father is older than his mother.’

    b. Test Data:

        Tha   mi a’   sireadh   an   leabhair bhig
be-PRES-IND 1S PROG searching-VN ART book-G small-G
ghuirm.
blue-G

‘I am looking for the small blue book.’

(2) Baseline:

  a. Trained with:
     Gaelic sentence → English Sentence

     i. e.g.
        Tha a athair nas sine na a mhàthair.
        →
        His father is older than his mother.

  b. Test with:
     Tha mi a’ sireadh an leabhair bhig ghuirm.
     →
     Predicted_English_Translation$_j$

  c. Evaluation of the model:
     compare
     Predicted_English_Translation$_j$
     with
     ‘I am looking for the small blue book.’

(3) Gloss Treatment:

  a. Trained with:
     Gloss line → English Sentence

     i. e.g.
        be.pres 3sm.poss father comp old.cmpr comp 3sm.poss mother
        →
        His father is older than his mother.

b. Test with:
be-PRES-IND 1S PROG searching-VN ART book-G small-G blue-G
$\rightarrow$
Predicted_English_Translation$_j$

A   Supplemental Material

Round	Baseline	GLOSS	ParaPart	Para	Interleaving	Concat	Google Translation
0	17.29	18.39	32.64	25.42	13.67	15.42	22.09
1	16.42	18.00	32.28	25.32	12.49	14.31	25.38
2	15.29	16.02	29.94	20.72	11.01	15.38	23.72
3	15.97	20.22	31.18	22.22	12.33	14.18	23.21
4	17.79	19.02	32.83	24.27	12.56	18.63	22.31
5	16.73	15.53	31.11	24.55	12.13	14.89	23.41
6	17.11	18.00	32.19	27.03	11.55	15.16	24.53
7	16.37	20.08	33.52	25.34	12.78	15.20	22.78
8	15.93	15.82	30.93	24.24	12.43	15.50	25.67
9	16.99	15.93	34.35	25.96	11.65	15.72	23.42
Mean	16.59	17.70	32.10	24.51	12.26	15.44	23.65

Table 1: BLEU scores of the treatments: Ten rounds of repeated random sub-sampling validation are conducted. For each round, the same sets of IGTs are used. Each column is a treatment, and each row is a single round of repeated random sub-sampling validation. The last column is the scores of Google Translation. We used a free Google translation API (**?**) to translate the same set of test Gaelic sentences into English.

itemize